DOCUMENT RESUME

ED 150 330 CE 014 519

AUTHOR	Hillestad, Mildred
TITLE	Research: Process and Product. Delta Pi Epsilon Service Bulletin No. 1.
INSTITUTION	Delta Pi Epsilon, St. Peter, Minn.
PUB DATE	77
NOTE	146p.
EDRS PRICE	MF-$0.83 HC-$7.35 Plus Postage.
DESCRIPTORS	Analysis of Variance; *Data Collection; Experiments; *Literature Reviews; Measurement Techniques; *Questionnaires; *Research; *Research Design; *Research Methodology; Research Problems; Research Tools; Technical Reports; *Writing; Writing Skills

ABSTRACT

This bulletin is designed for the researcher, the doctoral or master's candidate, or teacher interested in and/or conducting research. Content is presented in five chapters. Chapter 1 (The Research Problem) discusses what research is, selecting a topic, defining and stating a problem, and identifying the type of study. Chapter 2 (The Literature Review) covers purposes of a literature review, scope of the literature search, systematizing the search, indexes to the literature, preparing and using a master bibliography, and reading the literature and making notes. Chapter 3 (Data Gathering) discusses planning the data-gathering activities, questionnaires, interviews, observations, task analysis, attitude measurement, Q-Sorts, and the Delphi technique. Chapter 4 (Experimental Research) covers designing experiments, controlling variables, and analyzing experimental data. The last chapter (Writing the Report) discusses introducing the study; need for the study; purpose of the study and statement of the problem; delimitations and limitations; definition of terms; review of related literature; describing the procedures, reporting the findings; discussing the findings; summarizing, concluding, and recommending; compiling supplementary pages; and writing and organizing the report. (EM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED150330

RESEARCH: PROCESS AND PRODUCT

Mildred Hillestad

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ΔΠE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CE 014 519

Delta Pi Epsilon
Service Bulletin No. 1

Preface

Delta Pi Epsilon launched the new planned publications program during the 1976-77 biennium with two series, one targeted at the researcher and the other, the classroom teacher.

The *Service Bulletin* series is designed for the researcher, the doctoral or master's candidate, or teacher interested in and/or conducting research. The *Rapid Reader* series is geared for the teacher to read and immediately take ideas into the classroom for enlivening and enriching the teaching of business subjects. Each publication in both series is written by a recognized authority in the field. Several are Delta Pi Epsilon Research Award or Robert E. Slaughter Research Award winners.

Delta Pi Epsilon is proud to present these new publications for use in classes such as methods of teaching, principles of business education, and research classes as well as for use of teachers who want to remain abreast of the latest teaching/learning strategies and research techniques.

Publications released during the 1976-77 biennium include:

Service Bulletin No. 1—*Research: Process and Product* by Mildred Hillestad, The Ohio State University, Columbus, Ohio

Service Bulletin No. 2—*Design and Conduct of Educational Surveys and Experiments* by Leonard J. West, Baruch College, The City University of New York, New York, New York

Rapid Reader No. 1—*Teaching General Business* by Betty Jean Brown, University of Tennessee, Knoxville, Tennessee

Rapid Reader No. 2—*Teaching Consumer Education* by David K. Graf, Northern Illinois University, DeKalb, Illinois

Information concerning the above publications and others to be issued in each series may be obtained from the National Office of the Fraternity.

©Copyright, 1977 by Delta Pi Epsilon. Printed in the U.S. of America

---

DPE NATIONAL EXECUTIVE BOARD:

National President: Wanda Blockhus, School of Business, San Jose State University, San Jose, CA 95192.

National Vice-President: Robert A. Ristau, College of Business, Eastern Michigan University, Ypsilanti, MI 48197.

National Secretary: John S. Nigro, North Haven High School, North Haven, CT 06473.

National Treasurer: Jeanne L. Holley, Box 24, Conner Hall, University of Mississippi, University, MS 38677.

National Historian: Robbie Woodall, Nashville State Technical Institute, Nashville, TN 37209.

National Past President: Gordon F. Culver, 300 Teachers College, University of Nebraska-Lincoln, Lincoln, NE 68588.

DPE NATIONAL PUBLICATIONS COMMITTEE:

Chairman: Richard J. Dallas, Santa Rosa Junior College, 1501 Mendocino Avenue, Santa Rosa, CA 95401.

Member: Charles R. Hopkins, 254 Peik Hall, University of Minnesota, Minneapolis, MN 55455.

Member: Oleen M. Henson, 203 Ritter Addition, Temple University, Philadelphia, PA 19122.

Member: Carmela C. Kingston, Division of Business, Trenton State College, Trenton, NJ 08625.

DPE NATIONAL OFFICE:

National Executive Secretary: Ellis J. Jones, Gustavus Adolphus College, St. Peter, MN 56082.

Table of Contents

Preface . . . . Inside front cover
Table of contents i

Chapter 1 The Research Problem 1

- What is research? 1
- Selecting a topic 2
  - Characteristics of a good topic 2
  - Sources of topics 4
- Defining and stating a problem 6
  - Criteria for good problem statements 6
  - Analyzing the problem 10
- Identifying the type of study 12
  - Classifying variables 13
  - Recognizing the types of studies 16
  - Recognizing legitimate conclusions 20

Chapter 2 The Literature Review 22

- Purposes of a literature review 22
- Scope of the literature search 23
- Systematizing the search 24
- Indexes to the literature 25
  - Periodical literature 25
  - Books 27
  - Government publications 27
  - Theses and dissertations 27
- Preparing and using a master bibliography 29
- Reading the literature and making notes 30

Chapter 3 Data Gathering 33

- Planning the data gathering activities 33
  - Selecting the data gathering instruments 34
  - Advantages and disadvantages of survey data-gathering techniques . 38
- Questionnaires 40
  - General considerations in using questionnaires 40
  - Developing a good questionnaire 42
  - Writing the cover letter and follow-up reminders 60
- Interviews 60
- Observations 61
- Task analysis 63
- Attitude measurement 64
- Q-Sorts 68
- The Delphi Technique 70

Table of Contents (continued)

Chapter 4 Experimental Research 74

Designing experiments 74
Controlling variables 83
Internal and external validity 84
Analyzing experimental data 87
Analysis of variance 88

Chapter 5 Writing the Report 96

Introducing the study. 96
Need for the study 97
Purpose of the study and statement of the problem 98
Delimitations and limitations 101
Definition of terms. 103
Review of related literature 103
Describing the procedures 108
Reporting the findings 113
Discussing the findings. 126
Summarizing, concluding, and recommending. 128
Compiling supplementary pages. 131
Writing and organizing the report 132

References 137

Chapter 1

The Research Problem

Almost every graduate program in business education requires some type of research experience—either conducting research or reading and evaluating research done by others. Either of these experiences requires knowledge of the research process including the product, the research report. Without knowledge of the research process, a person cannot intelligently evaluate the research of others nor can he or she conduct defensible research.

What is a research study? Why do we have to conduct research or read research? How do we know what to study and how do we go about doing it? How do we know whether or not a study is well done? Such are the questions students sometimes ponder. This Service Bulletin addresses these and other questions regarding the process and product of research.

What Is Research?

This question can be answered with a dictionary definition: "Studious inquiry...critical and exhaustive investigation or experimentation having for its aim...the revision of accepted conclusions...in the light of newly discovered facts...." (Webster, 1971) Books on research methodology generally agree with this definition. Mouly's (1970) definition of research is clear and complete: "Research is best conceived as the process of arriving at dependable solutions to problems through the planned and systematic collection, analysis, and interpretation of data." (p. 12) The key words in this definition are **planned and systematic.**

Using research techniques to solve problems is more efficient than either trial-and-error attempts or "shotgun" approaches, since one directs all efforts toward a particular goal based on a sound hypothetical solution. Thus, many directionless, nonproductive paths may be avoided by a well-planned research effort. The systematic study of problems in business education advances knowledge in the field by testing theories establishing relationships among variables and determining how theories may be applied in classroom situations. Research can help develop curriculum, it can provide information about the nature of students and teachers and their behavior, or it can give us insight into the learning process.

The core of the research effort, though, is having a good problem that asks the right kinds of questions in the right way. However, mere collection of facts, whether from reference books, from readings in a library, from historical documents, or from a questionnaire is not research (Mouly, 1970, p. 13) unless the information derived from the analysis of the data solves a problem. As Van Dalen (1973) says in writing about such activity and data gathering by novices without first establishing an actual problem:

> The end result of their furious figuring and fact finding is a formless, frustrating fund of data. Having never decided what to solve, they roam rudderless in a sea of facts with their chartless voyage committed to failure. (p. 105)

Selecting a Topic

How, then, do you as a graduate student decide on a research topic, and what is a suitable problem? These processes are discussed in the sections that follow.

Characteristics of a Good Topic

Several characteristics of a problem should be taken into account in searching and deciding on a research topic:

1. The problem must be of interest. The topic should be one of great interest to you in achieving your professional goals. It should be interesting enough that curiosity will keep you motivated to search for the answer to your problem, even though considerable time and effort are demanded of you as a researcher. Occasionally the project may not go as smoothly as you would like, but being vitally interested in the topic will keep you going in spite of seeming obstacles until the project is completed.

This interest in the topic, however, should be that of curiosity; that is, it should not be a biased interest. The researcher should be interested in the outcomes regardless of whether they support any preconceived ideas or not. That is not to say, though, that the researcher has no idea of what the outcomes might be. Without having some notion of possible outcomes, you don't have a problem. In fact, an important step in the scientific method of problem solving is to pose probable solutions.

Not only should the topic be of interest to you, but it should also be interesting to others in the field in which you are working. Findings should, if possible, be applicable to more than your own school situation in one class. More will be said about this later.

2. The problem should be circumscribed. The topic selected for study should be limited enough to be feasible—in terms of time, money, availability of facilities and tools, knowledge and background of the researcher, and data that are collectible.

Will collection of the data take too long? be too expensive? Are adequate instruments available to collect the data: attitudes, motivation, socioeconomic status, physical manifestations? For instance, if you want to study the quality of decisions made by students completing their study of office procedures in different types of programs, a valid, reliable measuring instrument to measure decision-making ability would perhaps have to be developed. This in itself would be a long-term task, a difficult one, requiring much time and a great amount of technical background in measurement theory. Considerable cost might be involved also in gathering data from an appropriate sample representative of different types of jobs in different sizes and types of business. One may need to observe and talk with employees about the circumstances surrounding the decisions made, how the individual arrived at the particular decision, and the consequences of the decision.

Also, developing such a measuring device requires much technical knowledge about establishing validity of such an instrument as well as in determining its reliability. An inexperienced researcher may be better off settling for a comparison of achievement in the office skills and knowledges of office procedures gained by the students, for which some measures may already be available or more easily constructed.

3. The problem should possess a degree of originality. The problem should possess some degree of "originality" in that it is not an exact replication of a study that has already been done. Many studies need to be redone, but not in exactly the same way the first investigations were done. A restudy of a problem may be done to remedy any weaknesses in research procedures used by a previous investigator. New and better measuring instruments and methods may be available that were not in existence at the time of the original study. Some of the variables used by others may not have been carefully enough controlled so their findings are contradictory. The analysis of the data may not have been appropriate for the design of the study to answer the questions posed by the investigator. Consider, for instance, a study being carried out in one school that compares two sets of teaching materials but has two teachers each using a different set of materials. If one class makes higher achievement scores than the other, is the difference attributable to that teacher's general effectiveness or to the materials used in that class?

Studies of shorthand classes have been carried out using tests that students opted to transcribe as measures of achievement rather than comparing everyone on the same test given at the same rate and counting the actual number of words the students could transcribe. In basic business and bookkeeping research, outcomes have been measured by tests prepared by the investigator for which neither validity nor reliability was established (or at least not reported). Some typewriting studies in which net words a minute were the basis of comparison bear repeating, using separate speed and accuracy scores or combining them in multivariate analysis. Experiments have been reported in which no analysis was done to determine equality of ability of the groups at the beginning of the experiment or in which no attempt was made to select the students randomly or to assign the experimental variables at random to the groups. These problems in planning research studies will be discussed more fully in subsequent chapters.

4. The problem should be forward looking. Van Dalen (1973) adds another specification to the list of elements to consider in selecting a research topic: The topic of your research should, if possible, lead to further research, either directly on the topic itself or on some related matter closely allied with the topic or that emerges as the result of the investigation. Good research reports usually include among the recommendations suggestions for further research. A good researcher learns much from doing his research, and one question seems to lead to another. In fact, sometimes you may get the feeling that research raises more questions than it answers!

Sources of Topics

For many novice researchers, one of the most perplexing steps in the whole research project is HOW to go about selecting a suitable topic. Where do the problems come from? Where do researchers get their ideas, especially for the problems that meet the criteria of a good problem? The answer to that question is, "Anywhere!" However, having an alert, open, questioning mind from the start of your graduate work in classes—in reading, in talking with others—is one of the greatest aids you can have in finding a topic for research. However, the graduate student can take several actions to help in this selection process:

1. Do exploratory reading. Read everything you can: books, dissertations, master's studies, journal articles, not only in business education but also on any educational topic or in any content area. But in doing this reading, develop a "healthy skepticism" about almost everything a writer says. If someone makes a claim about the efficacy of some "new" classroom procedure, for example, but hasn't substantiated it adequately, ask, "On what basis is the statement made? What is behind it? What is the evidence? Does it, could it apply in one of my business classes?"

Reading outside the field of business education can bring to mind some interesting problems. Knowing what researchers in the field of reading are doing can lead to many promising ideas for research topics in teaching shorthand, for example. Such knowledge can bring up questions about effective ways to help students in business classes who have reading problems. Physical education literature provides ideas for the application of psychomotor principles in the teaching of typewriting. Reading about the teaching of social studies can be the source of ideas for research in teaching consumer economics or basic business courses.

Ideas about the application of educational and psychological theory to teaching business subjects may come from reading such journals as the *Journal of Educational Psychology, American Educational Research Journal,* the *Personnel and Guidance Journal,* and many others similar to these. Books and articles on the current res[illegible]ive learning theory are most fruitful sources of ideas for research in business education. However, one must "build the bridge" from the generalizations in these journals to business education, constantly keeping in mind the question

"How can this apply to teaching in my area of specialization in business education?"

Reading reports of completed research is another valuable source of ideas for research problems. In most good research reports, the investigator makes several recommendations for further research at the end of the report. Reading the entire report, however, should also raise questions, since the study may not have been done as carefully as it should have been to answer the questions posed. The procedures may not have incorporated all the elements that should have been considered in solving the problem, and these facets of the problem may still need to be investigated. Thus to use previously completed research as a source of problems, keep in mind such questions as these:

1. Were the procedures complete and correct?
2. Were appropriate and adequate measuring instruments used?
3. Do the findings apply to only one particular level or one type of student? Would the procedures produce the same results with students of a different age drawn from a different population?
4. Were appropriate statistical analyses made?
5. What was the theory underlying the research problem? How would that theory apply in a different kind of class?

To illustrate the kinds of inquiry that may arise from Question 5, consider the studies that have been done applying the principles of distribution of practice to typewriting classes, investigating whether as much can be accomplished in college typewriting classes meeting three periods a week as in classes that meet five periods a week. What about the application of this principle to shorthand or to office procedures? Would students achieve more in shorthand, for instance, in classes held for 20 or 25 minutes in the morning and another similar period in the afternoon than in classes that meet for one 50-minute period daily? Would the outcomes of office education programs organized in blocks of time of three or four periods daily and integrating all business courses be different from programs in which separate classes are taught in each of the subject matter areas that students need for preparation for employment?

Another helpful publication to consult for research ideas is Research Bulletin No. 3, *Needed Research in Business Education,* published by Delta Pi Epsilon (Hopkins, 1972). Again, as you read through the bulletin, ask questions about whether each of the ideas listed applies to other grade levels or classes.

2. Examine current practice. Examine classroom practices reported in the journals and those carried out in your own classroom. Ask yourself, "Why do my students seem not to understand this concept? If I approached it differently, or if different teaching materials were used, would the students learn more than they do now? Would that procedure recommended in that article or by that methods teacher really work in my classes?"

Question every suggestion made in graduate courses by the instructor and by other graduate students. Some of the "far out" ideas should reach your open mind, leading you to think about the theories underlying them. If it has a sound theoretical base, that far-out idea just might make a good research problem.

3. **Talk with others about the teaching/learning process.** As others talk with you about classroom problems, think about how what they say may apply to business education courses. A teacher of a graduate class might think aloud about the possibility of using a newspaper as reading material instead of the traditional textbook. What about a basic business or consumer economics class? Could one use the *Wall Street Journal* or perhaps a journal like *Money* or *Changing Times* instead of a textbook? Would the students learn the same things? Would they learn more than in traditional classes? Would they be more interested and motivated?

Defining and Stating a Problem

Once you have found a good problem area, it needs to be narrowed down to a specific problem. Novice researchers tend to want to tackle a problem that is too broad in scope, that encompasses too much. Instead of looking for the organizational patterns of office procedures instruction in the United States (which we shall see later is not a research problem but rather is merely a census-taking project), comparing the outcomes of office procedures programs organized in two or three different ways (traditional, cooperative, and block programs, for instance) would narrow the problem somewhat. Someone might suggest confining the study to schools within a restricted geographic area; this would reduce the geographic scope of the problem, but the **problem itself** is still not delimited. To delimit this problem, that is, to cut it down, questions might be raised regarding specific outcomes in these classes that should be considered—business knowledges, decision making, office procedures knowledges, basic skills development, attitudes, perceptions of office work. Thus, by asking a question such as "What is the extent of knowledge of office procedures gained by students in three types of office procedures programs: cooperative, integrated block programs, and traditional programs?" we have a delimited **problem**. Or a comparison might be made of the perceptions of office work held by students in the various kinds of office procedures programs. These, then are two delimited problems, each concerned with a relationship between two variables for which data is collectible.

Suppose that you now have an idea of a topic for research, one that has been reduced to a specific problem. How do you know whether or not you have a good research problem?

Criteria for Good Problem Statements

According to both Tuckman (1972, p. 20) and Resta and Baker (1973, pp. 105-108), a good problem statement must meet four criteria:

1. The problem must ask a definite question.
2. It should be concerned with the relationship between two or more variables.
3. It should be concerned with variables for which the researcher can collect data.
4. It should contain no value statements.

Asks a question. A good research problem asks a definite question. Tuckman (1972, p. 24) suggests three elements of educational situations about which questions may be raised in order to delineate a problem: available inputs, instructional activities and organization, and anticipated outcomes. Facets of these three elements may be combined in many ways to derive questions. Inputs include such things as students, teachers, attitudes, the job market, and community standards. Instructional activities and organization include elements such as materials, curriculum, programs, student selection, and teacher-learner relationships. Outcomes may be changes in attitudes or knowledges, decision-making ability, meeting needs of business and society, or meeting needs of students.

A beginning researcher may make a statement to the effect that he or she would like to study reading problems in basic business classes. At this point, an advisor is likely to respond with some questions: What do you want to know about reading problems in basic business classes? Do you want to know how well students read when they enter class? whether or not the textbooks are of a reading level suitable for your class? whether or not teachers are doing anything about reading problems? what might be done to help your students improve their reading skills?

Let us say that you have noticed that your students in basic business classes seem to have problems understanding what they read. You wish to study the matter, but first you need to ask a question about the situation to which you can seek an answer—you can't just "study the matter." Perhaps you think about the difficulty level of the books you are using. Should you find another type of reading material for the students? Are there some remedial reading procedures that you as a business teacher could use in your class? After some preliminary investigation (reading journals and talking with the reading specialist in your school) you decide to try to help the students overcome their reading problems. So you now have a question: What will happen if I introduce remedial reading procedures in my basic business classes? and you are on your way to a research problem.

You may have read that the use of the Cloze technique has helped some young people improve their reading skills. You may have read that inserting questions periodically throughout the material to be read helps students focus their reading on something specific. Now your original question can be stated in a more specific way: What is the effect on achievement in basic business classes of interspersing questions into the textbook reading material and of using the Cloze technique? You now have a specific question concerning the relationship between variables, for which data are collectible, and which contains no value statements.

Involves a relationship. Some problem statements are concerned with relationships that are not too obvious, and others really don't have relationships in them at all. Which of the following questions are concerned with relationships between variables; and if no relationship is involved, how could you reframe the question to introduce a relationship?

1. What is the effect of listening training on the ability to transcribe mailable letters?
2. What is the relationship between students' socioeconomic status and enrollment in consumer economics classes?
3. What are the patterns of administrative structure of joint (area) vocational schools in the United States?
4. What kinds of office tasks are performed by office occupations students one year after graduation from school?
5. What are the attitudes of transcriptionists toward their jobs in word-processing departments?

Questions 1 and 2 are concerned with relationships. In Question 1, one variable is the presence or absence of a listening-training program, and the other is the number of mailable letters. Question 2 looks at the relationship between the number of students enrolled and their socioeconomic status. Clues to relationships are words like **related, relationship**, or **effect**.

Question 3 is an example of a fact-gathering project that some researchers might say is not a problem. The person asking that question perhaps had in mind finding out whether schools of different sizes or perhaps in different parts of the country have different administrative structures. Even if the question were reworded to read "Do administrative structures of vocational schools vary according to size (location) of school?" or "Does the administrative structure of large vocational schools differ from the administrative structure of small vocational schools?" it still is essentially a fact-gathering project even though it now contains a relationship. However, the answer to this question may be helpful in updating a course in administration and supervision of vocational education.

Question 4 might be reframed in terms of a relationship by asking, "Do the office tasks performed by office workers a year after their graduation differ according to the type of office education program in which they were enrolled?" Unless the researcher working on Question 5 introduces some element regarding size or type of business, extent or type of word-processing training, the types of work assignments, or perhaps personality characteristics of the employees, the question is not concerned with a relationship between variables.

Requires collectible data. Data in all of the problems in the last section are collectible, the next important criterion of a good problem statement. Data may sometimes seem to be collectible, but no really good instrument is available that will provide **valid** data. On the other hand, the investigator may be unaware of instruments that may be available for measuring certain variables. Before giving up a potential research topic, the researcher should consult appropriate sources regarding the availability of suitable instruments. Specific suggestions about such sources will be made in the next chapter.

Which of these problem statements contain variables for which data are collectible:

1. Does an extended student-teaching experience result in greater success as a teacher?
2. Are students in block-time office education programs more highly motivated toward learning than students in traditional programs?
3. What are the implications of the job experiences of teachers for teaching students for the office of tomorrow?
4. What are the effects of pacing on transcription rates of second-semester shorthand students?
5. Do we need to return to teaching the basic 3 R's to produce better office employees?

The data for Question 4 obviously are collectible. We can report the number of words transcribed in a given time, or we can report the length of time required to complete the transcription problems. How to get data for the other questions may be less obvious, however, or they may be uncollectible. Can you define teaching success in Question 1? If so, it then perhaps can be measured, but how? Some might say with supervisors' or principals' ratings, but ratings in general have low validity and reliability. Others may claim that teaching success is reflected in the number of students who can "pass" certain achievement tests. Little agreement has been reached on the definition of a good teacher in spite of many years' research on the problem.

Scales are available to measure student motivation and attitudes toward school, so data may be collectible for Question 2. One can collect data about teachers' work experience for Question 3, but the problem arises of classifying and quantifying such data. What kind of experience? How can the kinds of experiences be weighted in relation to the amount? Is one more important than the other? Also in Question 3, what is meant by "implications"? As it stands, the question has not been refined into a research question. Question 5 poses the problem, of course, of what is meant by "better." That term can mean many things, each according to the reader's value system. Are "better" employees those who are more diligent? more productive? more loyal to the firm? get along with others? willing to work overtime?

Contains no value statements. Question 5 above, violates the fourth criterion of a good problem statement, because it contains a "value" term, "better." Other examples of value words to be avoided in problem statements are **bad, good,** or **should.** Which of the following questions are value statements and which are adequate problem statements?

1. Is punishment in a typing class bad?
2. What are the effects of teaching mode on achievement test scores in business law classes?
3. Is team-teaching better than a single teacher in a block-time office education program?
4. Do students who have had a business ethics course have higher moral standards than those who have not?
5. Do tenth-grade girls have better attitudes toward studying money management than tenth-grade boys?

What is meant by **bad** in Question 1? Bad for whom, students or teachers? How is badness measured? Is it a tense emotional atmosphere in the classroom? Is it more errors on timings or fewer usable letters? Is it unfavorable attitudes toward the class? The third and fifth questions bring up the word **better** again. In Question 3, what is better? knowledges? attitudes? work habits? initiative? Also, what does **single** mean, one teacher or an unmarried one? What is a **better attitude** in Question 5? Attitudes that agree with what the teacher has said in class? Accepting anything printed in the textbook? Changing "better" to "more favorable" as measured by a given attitude scale (valid and reliable) would make this question an acceptable one. In Question 4, **moral standards** itself is a value-laden term, its interpretation depending upon an individual's point of view and value system. Also, how are moral standards measured?

Analyzing the Problem

Posing an adequate research question is a good beginning step in planning a study, but the problem must be analyzed further, breaking it into the elements needed to solve the problem. This will entail asking a series of questions that will tell you not only what data are needed but will also give direction to how you go about solving the problem.

Specify the purpose of the study. The first step in analyzing the problem is to determine specifically the **use** to be made of the findings of the study. Is it a study that could result in curriculum or course revision? Is it a study that provides information about teaching methodology? Perhaps one is interested in providing teachers with information about employment standards in the community.

Ask the necessary questions. Once the researcher determines the specific purposes of the study, the questions that need to be answered in order to solve the problem must be formulated. By way of illustration, consider the teacher of an introduction to business class who feels that the students are uninterested in the textbook reading assignments. The textbook is read only perfunctorily, if at all. The teacher, feeling that the students will learn more if they read more than they do now, wonders how to get the class interested enough to read what is assigned. The teacher recalls that when students bring articles from newspapers or magazines, interest in the class seems to be greater. This observation leads the teacher to an idea: "What would happen if reading assignments were all from such articles instead of from the textbook? Maybe the students would learn more and have more positive attitudes toward the class." The teacher has also noticed that the more independent students seem particularly interested in the special reading materials, more so than less secure students. Perhaps the teacher notices that only the brighter students tend to want to read these materials or that students who usually do little reading seem more interested in class whenever the articles are brought in. If the students are more interested, will they learn more than they do now?

Let us help this teacher formulate a research problem from this situation. The teacher's purpose for doing such a study of differentiated reading assignments is to improve instruction in the introduction to business class, resulting in greater achievement on the part of the students. The basic problem statement then becomes:

> What is the effect of different types of reading assignments on the outcomes of an introduction to business class?

But this question is still too general. It needs to be delimited and terms need to be operationally defined before subquestions can be asked. Two terms especially, **outcomes** and **types of reading assignments**, need to be specified. Outcomes might include knowledges, attitudes, decision-making, amount of unassigned outside reading done, etc. Studying all of these may prove to be too large a task, so a decision must be made about what specifically to study. These decisions **delimit the problem.** The **delimitations** are those elements that might logically, normally have been part of the study, but which the researcher elects to exclude for good reason. If the teacher seems to be most interested in achievement of the students (in terms of knowledges), the decision may be made not to investigate attitudes toward the class in relation to the type of reading assignments made; thus, in excluding attitudes from the study, the problem is delimited. Delimitations pertain only to the problem itself and not to the population or sample used in the study.

A further decision is the specification of reading assignments. Since the teacher has reason to believe that newspaper and magazine articles hold the students' attention, the teacher decides to compare the effect of reading articles with reading textbooks. The problem then is specific, and it meets the criteria of good problem statements:

> What is the effect on achievement in an introduction to business class of reading assignments made in textbooks and those in magazine articles?

The statement (question) concerns a relationship (effect on) between two variables (type of assignment and achievement); it contains no value statements, and the required data are collectible.

The problem statement leads the teacher to wonder whether certain kinds of reading assignments benefit some students more than others, so we can specify more clearly groups of students about whom we are concerned by asking these subquestions:

1. Do boys and girls learn different amounts in an introduction to business class according to the type of reading assignment they have had?
2. Do students of different abilities learn different amounts according to the type of reading assignment?
3. Do students of different personality structures learn different amounts according to the type of reading assignment?

Or the questions may be framed another way to avoid repetitious questions:

Does the type of reading assignment affect the amount learned in an introduction to business class among students classified into the following groups:

Boys and girls
High, average, or low scholastic aptitude
Aggressive and submissive personality

Specify the needed data. The teacher must now consider the feasibility of collecting the data required for each of the subquestions and decide which aspects of the problem can be studied within a reasonable time at not too great a cost. The teacher decides that since ability scores are in the students' files, they are immediately available at no cost; but measuring personality traits is another matter, and the decision is made not to pursue that factor at this time.

Identifying the Type of Study

The teacher now ponders the question of how to find out whether or not these speculations about learning and reading assignments are true. Would asking other basic business teachers about it tell us what the best procedure is for different types of students? The teachers may or may not have tried differentiated reading assignments, but they may think it sounds like a good idea and will agree that it could be effective. But does such information really tell us whether or not one type of reading assignment is more effective than another? What we find out is only what those teachers **think** and not what is most effective. As George Bernard Shaw is reputed to have said, "If 50 million people say a foolish thing, it is still a foolish thing."

Instead, this research problem calls for comparing the achievement of groups of students whose reading assignments have been either from newspapers and magazines or from textbooks. But this raises the question of how achievement must be measured. Teachers, of course, have tests for their classes, but should achievement be measured by a test that the teacher makes up, one the publisher furnishes with the textbook, or one from another study? Whichever test the teacher selects must be reliable and valid for the particular purpose for which it is being used. Teachers probably do not know how reliable their tests are, nor do they know how valid they are Determining reliability and validity and developing a suitable test is a technical, time-consuming project; so our teacher needs another source of tests. A publisher's test? If validity data are given by the publisher, along with data about reliability, such a test may be suitable, but such information may not be available for tests that accompany textbooks.

Our teacher knows about a test prepared for a study concerning a unit in money management. The test was carefully developed, with the researcher having determined that it was valid for a group similar to the teacher's, using the same textbook as our teacher is using. Reliability was relatively high. But the test covers only the money management unit. Thus if the decision is to use that test, the research question will have to be revised once more. Here we have an example of the interdependence of delineation of the problem, the research methodology to use, and the collection of the data.

Classifying Variables

To understand further the types of research that you are reading or that you might be thinking of carrying out, you must look more closely at the types of variables involved. The types of variables in the study are what determines the type of study; the statistical analysis does not. The choice and manipulation of the variables and the particular relationship between them determine the type of statistical procedure that must be used; the statistical procedures do not determine the type of study. Thus, to determine the appropriate approach to studying a problem and to identify the type of problem being proposed, one must recognize the characteristics of the variables involved.

Independent and dependent variables. First, one must know the difference between dependent and independent variables, or variates, as Sullivan (1972, pp. 4-5) labels them. The dependent variable (or criterion variable) is the element that may or may not change because of the treatment (procedures, materials, teachers, etc.) to which the subjects are exposed. What happens to the criterion variable "depends" on the treatment given to the students (subjects) under study. In our teacher's problem of what happens as a result of using special reading assignments instead of textbooks, the achievement scores constitute the **criterion** or **dependent** variable.

The **variates**, the **independent variables** (procedures, materials, physical conditions, etc.), are those whose effects are being studied; they affect the criterion variable directly or are related in some other way to the criterion. These factors do not change because of the procedures being used. In our problem, use of special reading assignments may affect the amount the students learn; therefore, the types of reading assigned are independent variables (variates).

As a subproblem, the teacher is interested in the relationship between the students' abilities and the type of reading assignment they are given and how these together affect learning. Achievement is still the criterion, but we now have a second variate, student ability. It is independent of the procedures being used—the student's scholastic aptitude does not change as a result of reading either periodical articles or a textbook. The researcher chooses to classify students according to their ability before the study begins, so ability is an independent variable.

Thus, the use made of a variable and not the variable itself determines whether it is an independent or dependent variable. For instance, achievement scores, the criterion in this study, could serve as the basis for classifying students for another study to determine whether or not the attitudes toward consumer protection laws of individuals with greater knowledge of basic business differ from the attitudes of students whose scores were among the middle and lowest thirds of their class. So what was a criterion variable in one study may be an independent variable in another.

Before you read further, test yourself to see whether you can identify types of variables. The variables listed here may be dependent, independent, or either. Decide which each one is, and check your responses in the discussion below:

1. Sex of the students
2. Students' reading rates
3. Absence rate
4. Score on the California Test of Mental Maturity
5. Class size
6. Participation in FBLA
7. Score on the Test of Economic Understanding

Variables 1 and 4 are not likely to change as a result of a treatment. No classroom procedure prescribed by a researcher will change these characteristics of a student, who is either a boy or a girl with a certain amount of intelligence (ability) when the study begins. The other variables may be either dependent or independent, although Variables 3 and 7 are more likely to be dependent (criterion) variables. Absence rates may go down (or increase) if a new procedure is introduced, and scores on the Test of Economic Understanding can be affected by some new teaching procedure or materials, or it may be related to a given level of ability of the student. However, these scores could be the basis for classifying students for another study, and thus be independent variables.

Participation in FBLA could be the result of different types of recruitment activities, making FBLA membership the criterion. On the other hand, characteristics or achievement might be compared of students who do and those who do not participate in FBLA, making FBLA membership a variate (independent variable).

Thus, a variate (independent variable) is a variable that the researcher "manipulates" or one that serves as a basis of classification of the research subjects **prior** to the beginning of the study. However, identification of variates and criterion variables is usually easier when they occur in problem questions. Which is the criterion in these problem situations?

1. How does
    teaching business arithmetic on calculators
  affect
    arithmetic achievement?
2. Is
    teacher estimate of student ability
  related to
    grades assigned to students' work?
3. Is there a relationship between
    length of office simulation
  and
    students' perceptions of selected office jobs?

Manipulable and nonmanipulable variables. Another even more useful way to classify variables according to Sullivan (1972, pp. 3-9) is to determine whether or not they can be manipulated by the researcher. If one can identify manipulable and nonmanipulable variables and recognize whether or not they were manipulated, one can determine the type of study one is dealing with, which, in turn, determines the types of conclusions that can legitimately be drawn from the findings.

Sullivan (1972) defines manipulable variables in two ways:

1. Those elements in a study that can be changed or controlled by the researcher, and
2. Those variables (pre-existing) that may differ in degree, amount, or kind to which the subjects under study can be randomly assigned. (p. 7)

In our example of a study concerned with the use of different reading assignments in a money-management unit, the type of reading assignment is the first type of manipulable variable. Reading assignments can be selected and arranged in whatever way the teacher desires, and students can be randomly assigned particular types of reading materials.

Examples of the second type of variable (characteristics pre-existing in the situation to which the students may be randomly assigned) may be the sex, age, or years of office experience of the teachers; the type of office education programs that may be offered in a school. Students can be assigned to either a male or female instructor, or they could be assigned randomly to older or younger teachers; or if the school has more than one type of office education program, students conceivably (but perhaps not practically) could be randomly enrolled in the different programs.

Nonmanipulable variables are characteristics of the subjects of the study to which they cannot be assigned, conditions that cannot be changed by the researcher. Examples of such variables might be the parents' occupations, the number of parents in the student's home, the student's age, aptitudes, or brain hemisphere dominance. None of these variables can be changed; the student has them at the beginning of the study and the experimental procedures will not change them.

Applying either of the definitions of manipulable variables, decide which of these variables are not manipulable:

1. Number of siblings.
2. Frequency of club meetings
3. Number of pages read as homework (in shorthand)
4. Degrees held by the teacher
5. Teacher's age

Deciding whether or not a variable is manipulable may not always be as easy as it might seem. Quite easily, however, the second and third variables may be recognized as manipulable and the first one as nonmanipulable. However, depending on the situation, Variables 4 and 5 may or may not be manipulable. If students could be randomly assigned to teachers of different ages, or to teachers holding different academic degrees, those variables are manipulable. However, if the teachers themselves are to be studied, their ages or degrees may be nonmanipulable, as in this research question:

> Do teachers between the ages of 35 and 50 who are just beginning their teaching careers (individuals, for instance, who have returned to school to pursue a "second career") use teaching practices different from those used by younger beginning teachers who have had no previous career?

Degrees held by teachers is a nonmanipulable variable if a researcher is trying to determine if teachers with master's degrees use more "thought-starter" questions than do teachers who have a bachelor's degree. In these two situations, the teachers cannot be assigned to degree designations or to age groups; they are selected for study **because** they already have the particular degree or are a particular age.

Recognizing the Types of Studies

Researchers often classify studies according to the methodology and/or the type of statistical analysis done. However, Sullivan (1972) uses a simple categorization based on manipulation or nonmanipulation of variables. It includes three categories: status, associational, and experimental studies. Characteristics of each of these types of studies will be discussed and illustrated.

Status studies. Studies in which the variate (independent variable) is not manipulable are **status studies.** If a researcher has as a problem, "Do 10th grade boys and girls differ in their perceptions of the American business system?" the study involved is a status study. The variate (sex of the subjects) is nonmanipulable. Status studies often are surveys of some sort. They are often referred to as normative-survey studies in research methodology textbooks. However, because of other factors that could be introduced by the subproblems, the statistical analysis may be quite sophisticated and may even be the same as that used in experimental studies. The fact that the variate is not manipulable, however, makes it a status study because the sex of the students does not change nor can the students be randomly assigned to a particular group. They are in a particular group (male or female), because they are either boys or girls. The study remains a status study regardless of the complexity of the statistics used in the analysis of the data.

Which of these problems are status studies?

1. Do students who have graduated from high school with one year of bookkeeping and those who had two years of bookkeeping/accounting instruction perform different types of accounting/bookkeeping tasks on the job six months after graduation?
2. Do seniors in college with social studies majors, business administration majors, or business education majors hold different attitudes toward business and government?
3. Are students' career choices related to parents' educational levels?
4. What is the relationship between amount of time spent on practice sets in high school bookkeeping classes and scores on the CPAI Aptitude Test?
5. Is dropout rate from block-time office occupations programs related to size of school?

All but Question 4 involve status studies. The students in Questions 1, 2, and 3 would be selected because of the amount of bookkeeping/accounting instruction they had had, because of their major, or because of their parents' educational background; schools would be selected for Question 5 because of their size. In Question 4, the amount of time spent on practice sets could be manipulated, so this question does not involve a status study.

Let us consider another problem question: "Do basic business teachers in large school systems have different attitudes toward use of games and simulations in basic business classes than similar teachers in smaller school districts?" This is a straightforward status study, since teachers are not assigned randomly to school districts of varying size. However, if the researcher feels that perhaps a greater proportion of the teachers in large districts will have master's degrees than teachers in smaller school systems do, a subproblem could be formulated: "Are the academic degrees possessed by basic business teachers related to their attitudes toward the use of games and simulations in teaching basic business courses?" Other subproblems could be introduced also, such as the sex of the teacher in relation to use of games and simulations in class; but each variate introduced increases the complexity of the statistical analysis and makes it similar to that used for some experiments. However, since teachers are selected because they are in a particular school and cannot be randomly assigned to the subgroups, it is still a status study.

Associational studies. A second category is **associational studies** in which the variate is manipulable but is not manipulated. These studies usually involve two or more groups of subjects who already possess different amounts or degrees of a manipulable variable. A recent example of such an associational study is McLean's (1975), in which he studied the office-work preparation and the job satisfactoriness of graduates from three types of office-education programs. Rather than assigning the students to the three types of programs and evaluating differences two years hence, the subjects were selected **because** they had received their preparation for office work in one of three ways: model office, traditional program, or cooperative office education program. Thus, the variate under study could have been manipulated, but was not. Students could have been randomly assigned to the three programs (in very large systems, perhaps); but this would have been a rather impractical approach to the problem in terms of time and cost requirements. Rather, McLean chose to study students who had received their preparation in one of the three different ways, making it an associational study. Some research methodology books would classify these as **ex post facto** studies. Again, the statistical analysis in this study is similar to that used in experimental studies, but because the variate was not manipulated, it is an associational study.

Experimental studies. Had McLean been able to assign the students randomly to the type of office education program, his study would have been an **experimental study**. Thus, experimental studies are those in which the variates (independent variables) are manipulated and/or in which **subjects are randomly assigned to the treatments**. Only by looking at the procedures used by the researcher can one determine the type of study one has under consideration. In which of the following research situations is the researcher concerned with an associational or with an experimental study? Are any of the situations status studies? Which is the variate (was it manipulated?) and which is the criterion? Assume in each situation that adequate control procedures were followed and that suitable measuring instruments were available.

1. **Research question:** Is the amount of time spent on reading shorthand plates and homework notes in class related to the degree of dictation skill developed in beginning shorthand classes?

   Procedures: One teacher with two sections of beginning shorthand was asked to spend 10 minutes a day in one class section reading shorthand plates and 25 minutes on dictation practice. The other class spent 10 minutes daily on dictation and 25 minutes reading plates and homework notes. Reading and dictation materials were exactly the same, and theory presentations were conducted the same way in both classes. At the end of the first semester (after several intermediate tests had been given) reading rates and dictation skills were measured.

2. **Research question:** What is the effect on business math achievement scores of using calculators for all problem-solving exercises and of solving the problems with only paper and pencil?

   Procedures: One teacher teaches two classes in business math: the first-hour section meets in the office education laboratory, where a calculator is available for each student in the class. The second-hour section is assigned to a general classroom set up with tables but no calculators. The teacher compares the achievement of the two class sections.

3. **Research question:** Is the number of mailable letters transcribed by students related to students' scores on a business vocabulary test?

Procedures: The students were classified according to vocabulary test scores into high, middle, and low groups. This was done the first day of a fourth-semester transcription class. The students transcribed two letters daily for a period of four weeks. The achievement score for each student was the number of mailable letters transcribed.

4. **Research question:** What is the effect on achievement in a salesmanship class of videotaping students' sales presentations after they viewed videotaped model sales presentations by professional sales people or after discussion of sales techniques, with no videotaped model?

Procedures: A series of four videotapes of model sales presentations was shown to and discussed by a randomly selected half of a salesmanship class. The other half of the class discussed effective sales techniques for each of the products sold in the videotaped demonstrations, but this group was not shown the tapes. Each student in the class made one practice sales demonstration, which was critiqued by the class, and then was videotaped making a sales presentation of a product related to those in the tapes. Each student's videotaped presentation was rated by a panel of experts.

Of the four studies outlined, only Problems 1 and 4 may be considered true experiments. The students were assigned to the experimental treatments, and the variables were actually manipulated (deliberately made different for each group) by the researcher.

In Problem 2 the researcher did manipulate the variates, using two different approaches to teaching business math. However, students were not randomly assigned to the two class sections. Studies in which intact classes are assigned to the treatments are labeled by Campbell and Stanley (1966, p. 47) as quasi-experimental studies. Had the researcher been able to randomly assign students to the two teaching conditions, the study would have been a true experiment.

Problem 3 is not an experiment since the variate (vocabulary level of the students) is not manipulable. This is a status study that would be classified as an **ex post facto** study in many research textbooks. If the students had not been grouped into high, middle, and low vocabulary groups, but a coefficient of correlation had been calculated to measure the degree of relationship between vocabulary and transcription skill, this would be classified in many research textbooks as a correlational study.

In all of these problems other variates could have been introduced, making the statistical analysis in some cases quite complex; but if you understand the types of variates involved and whether or not they are manipulable, and whether they are or are not manipulated, you will be able to classify studies correctly. You will realize that the type of study determines the statistics to use and that the particular statistical analysis does not determine the type of study you are dealing with. You will also know whether certain conclusions may legitimately be made as you read or plan studies.

Recognizing legitimate conclusions

The conclusions that may be drawn from each of the three types of studies differ. Status and associational studies are studies of relationships, not of cause and effect. Only as a result of experimentation may statements of cause and effect be made.

Status studies often report only straightforward facts; for instance, the number of companies that have magnetic tape or magnetic card typewriters, or the kinds of training programs they provide for employees to learn to operate the equipment. However, if the sample of companies was divided according to size or type of business, the researcher would be concerned with relationships; that is, the relationship between type of company and number of magnetic typewriters or between the type (size) of company and the use made of such equipment. Although a status study may show, for instance, a relationship between the size of the company and the general dissatisfaction of the word processing employees with their jobs, one cannot legitimately say that the size of the company is the **cause** of the dissatisfaction. The actual cause of the dissatisfaction may be a third element not even considered in the study; for instance, routineness of the typing tasks may be the case in larger companies. Perhaps the work in large word-processing departments may be fractionated so much that all planning and setting up of the material is done by one person and the majority of the word-processing employees do nothing but play back the tapes. Thus, not the size of the company, but rather the monotony of the work may be the cause of employee dissatisfaction. Relationships are just that, relationships, not cause.

Knowing the type of inference you can make from a study is important in planning a study. If you wish to make statements about the **effect** some classroom procedure has on student achievement or attitude, you will conduct an experiment and not survey a thousand teachers to find out what they think. Only when the researcher controls the manipulation of the variates (or by randomly assigning individuals to the treatments) can statements be made that the outcomes were the result of (were caused by) the variate being studied.

In a study concerned with whether students in rural schools score differently on an economic understandings test than students in urban schools do, only statements concerning the relationship between the location of the school and economic understanding of the students may be made. Statements may be made about relationship, but not about cause. Whatever conclusions are allowable in associational studies may be made also in experimental studies, but the opposite is not true.

Which of the following statements of conclusion may be legitimately made only in experimental studies:

1. Developing good grammar and punctuation skills in machine transcription before learning shorthand transcription results in higher shorthand mailable transcription rates.

2. The type of office education program in which beginning employees received their office training has no effect on their scores on an office perceptions scale.
3. Little relationship exists between S.A.T. scores and straight-copy typing skill.
4. Use of calculators results in no greater achievement than paper-pencil methods in solving business math problems.
5. A greater proportion of students in rural high schools study bookkeeping than in large urban schools.

If you selected 1 and 4 as conclusions legitimate only in experimental studies, you would be right. The statements, that particular classroom practices result in higher achievement, are cause and effect statements and thus are permissible only in experimental studies. Conclusion 4 contains the word "results," a cause-and-effect statement, as does the first conclusion. Conclusion 2 may have been from an experimental study, and should be made only in experimental studies, since it contains the word "effect." Only when the researcher controls the manipulation of the variates can statements be made that outcomes were the result of (were caused by) the variate being studied.

Chapter 2

The Literature Review

One of the most important steps in delineating and carrying out a research project is searching the literature related to the problem you have in mind. A thorough review of the literature contributes in many ways to a research study.

Purposes of a Literature Review

Reading is done initially to help define and delimit the problem, to get ideas of how others have approached similar problems, to develop hypotheses, and to discover just what is known about the field. This reading should be broad, including opinion articles as well as studies. Opinion and philosophical articles are rich sources of research ideas, since many opinions expressed are conjecture and speculation about "what would happen if..." that need testing in a research situation. Many ideas come from these writings if you keep questioning the author's claims as you read and if you are not willing to accept what the author says without looking for the theoretical base or evidence that the ideas propounded have been compared with other ideas. Previously done research studies presenting recommendations for further research often are fruitful sources of problems and approaches to them. Broad reading at the beginning stages of the research process also acquaints the researcher with others who are investigating problems similar to your own area of interest. Tracing the work of these persons and reading the references listed in their bibliographies gives you a picture of what has been done before.

Once the problem has been stated and you are on the way with planning the investigation, the reading changes to more specific, more critical analytical reading than the initial exploratory reading. You then look for new approaches to the problem, to see what measures might have been used by others or what might be available. You may find work done with one group of subjects that might be applied to groups of other ages or with different backgrounds that you are interested in.

Knowing what others have done and found helps you avoid weaknesses of others' studies and can save you from "re-inventing the wheel." However, all reading must be done critically, observing both the strong and weak points in the research methodology. Too many researchers pay attention only to the findings rather than looking at the findings in conjunction with the samples, methodology, and controls used. Always ask as you read, "Did the findings really come from the variables under study, or might they be the result of other factors that the researcher paid no attention to?"

Scope of the Literature Search

How does one find the right things to read? How does one know that enough has been read? The broader the preliminary, exploratory reading one does, the better. The more widely you read, the more ideas you can get of what to do and what not to do. Read not only in your own field, but also read studies and articles closely related to your problem, its methodology and/or theoretical base. Reading outside the content field in which you are planning a study is especially important in establishing the theoretical base for the study, such as reading sociology or psychology or learning or personality theory or whatever forms the basis of the study being undertaken.

Reading the more current literature first has advantages. From such an approach you can develop a bibliography of what current investigators have read on the topic. Chances are that they have screened the literature and eliminated the less helpful studies. You may also find references to studies or articles not listed in the reference sources available to you. If certain studies seem to be cited repeatedly, they are perhaps the "classic" studies that should be included in your review of the literature.

The trend of the findings becomes apparent so that only the most "typical" and significant studies need be reviewed. Selective reading is particularly important if a great deal has been written about the topic under study. You cannot hope to cover the entire bibliography on many problems, so careful selection must be made of the studies that provide the appropriate framework and complete background for the study underway. This does not mean that the complete history of the discipline in which one is working needs to be traced; you need only to establish the framework for the particular aspect of the field being studied.

On the other hand, if the topic is a relatively new one (a new classroom procedure, for instance), one may read mostly in the subject-matter field or in an allied theoretical base for the study underway. For instance, for his study of the effect of using a constant rapid rate of dictation for all dictation practice in beginning shorthand, Boss (1967) had as theoretical base the micromolar theory of learning as discussed in an article in the *Harvard Educational Review* in which the experimental work applying the theory was done in arithmetic classes. At the time he did the study, Boss could locate no studies in shorthand using constant rates for speed development.*

Only original sources should be reviewed. Abstracts usually do not contain enough information about the procedures to allow a thorough evaluation of the study. They often lead to misinterpretation of the findings. Only by reading the original report in which the methodology and controls are detailed can one properly evaluate the findings. Neither can one depend on the review done by another researcher. The findings may be interpreted incorrectly or from the reviewer's particular bias. He or she may not have adequate background to evaluate the findings accurately, or may have given little attention to the procedures and concentrated only on the findings. A case in point is the author's study of factors contributing to the difficulty of shorthand dictation materials in which the findings clearly showed that the frequency of use of words in the dictation was a considerably more important factor than syllabic intensity. However, in one summary of research in business education, the only factor singled out as affecting difficulty of dictation matter was syllabic intensity.

---

*Notice the wording, "Boss could locate no studies..." One should never make the statement "There were no studies..." since someone may have completed a study somewhere that was not listed in the references consulted by the researcher.

Systematizing the Search

Reviewing the literature may seem to be a formidable task. How does one know what to read? Where does one find the things one wants to read? How does one go about it? Organization is the key word; approach the problem in a systematic manner.

List key words. Think of all the words that might be used for classifying the studies or articles of interest in indexes of periodical literature or of studies. Usually the classification schemes used in these bibliographic publications are given either at the front or back of the volume; and making a list of possible categories from those listed helps you systematize your literature search. Pay attention to "See" and "See also" notations in the indexes. They may provide valuable leads to other phases of the topic that you may not have thought of. Make a separate list of topics for each index you plan to consult.

Index: Business Education Index

Topic	1977	1976	1975	1974	1973	1972	1971	1970
Basic Business								
Consumer Ed.								
Equipment + Supplies								
Fast Learners								
Research								
Slow Learners								
Teaching Aids								

Figure 2-1. Check-off Sheet for Indexes Consulted

Search all appropriate indexes and references. Using your list of key words for the particular index you are consulting, follow it systematically for each issue of the index. Prepare a check-off sheet to record the topics and the particular volumes of the index you consulted (see Figure 2-1), and make a check mark by each topic consulted. Sometimes a particular volume of the index is missing or in use at the time you are preparing your bibliography, or you may be interrupted before you complete a volume, and the chart will show you at a glance which volumes and which topics you have already covered.

If you list the topics in alphabetic order, you can proceed through the volumes sytematically and efficiently and you will be less likely to overlook a topic.

Indexes to the Literature

Before you do any extensive reading, prepare a master bibliography of articles and studies that seem to be pertinent to your problem by consulting appropriate indexes and guides. Many specialized and general indexes to periodical literature are available, along with catalogs of books, pamphlets, government publications, and listings of research reports. Several of the more helpful and widely available indexes and bibliographic references are briefly described in the sections that follow.

Periodical Literature

1. *Business Education Index,* published annually by Delta Pi Epsilon (St. Peter, Minnesota: 1940-present). This index covers articles in business education periodicals as well as journals dealing with other business and office topics of an educational nature. It also lists masters' and doctors' theses and other independent research reports, as well as business education textbooks. Listings are alphabetic by subject and by author.

2. *Education Index* (New York: Wilson, 1929-present) lists periodical articles by subject; covers all fields of education.

3. *Business Periodicals Index* (New York: Wilson, 1958-present) formerly *Industrial Arts Index,* lists articles and books dealing with business and industrial topics. Alphabetic listings by subject.

4. *Social Science Index* and *Humanities Index* (New York: Wilson, April, 1974-present); formerly, 1920-1964, *The International Index to Periodicals,* from 1965 to April, 1974, when the *Social Science and Humanities Indexes* were combined into a single volume. These indexes list articles concerning social sciences, humanities and subjects of general interest. Alphabetic subject and author listings.

5. *Readers' Guide to Periodical Literature* (1900-present) indexes general articles from a wide range of magazines, alphabetically by subject and author.

6. *Psychological Abstracts* (American Psychological Association): Indexes some 800 journals, technical reports and monographs in psychology and allied fields including physiological, developmental, experimental, and general psychology as well as educational and applied psychology. It also indexes topics related to personality, social perceptions, and communications systems. Cumulative author and subject indexes are available for the years 1927-1974; each issue contains a subject and an author index.

ERIC Educational Documents and *ERIC Research in Education* are two other useful indexes for researchers in business education.

Only the more commonly used indexes have been listed here. Many more specialized ones are available in many libraries. You might consult the *Bibliographic Index* (1938-present) for information about bibliographies available in any subject area. This publication lists bibliographies published separately as books or pamphlets and also bibliographies given in other books and periodicals.

A good reference containing extensive bibliographies on educational subjects is the *Encyclopedia of Educational Research.* (Bibliographies such as these are listed in the *Bibliographic Index.)* Issues of the *Review of Educational Research* also contain rather complete bibliographies.

Should you need information about research done on tests, inventories, or scales, a most complete listing, including reviews of the instruments and bibliographies of research on each instrument, is given in the *Mental Measurement Yearbooks.* You may need to consult several editions of this reference to get a complete research background and evaluation of a particular scale or test, since the same instrument may be discussed in several different volumes over the years.

However, *Tests in Print* is the master index to all the volumes of the *Mental Measurements Yearbook* and should be consulted to find the particular volume(s) of the *Mental Measurements Yearbook* in which particular tests were reviewed. Also listed in *Tests in Print* are new instruments published since the last issue of the *Mental Measurements Yearbook.* It also includes bibliographies about the use made of the tests by researchers and of reviews of the instruments. *Tests in Print* itself does not contain reviews of the instruments it lists.

In addition, test reviews may be found in such periodicals as *Journal of Educational Measurement, Journal of Educational Research, Journal of Counseling Psychology,* and others. Several other volumes listing tests and scales, some published and some unpublished, may be available in your library. These volumes may be helpful in locating instruments to measure many other facets of the human character, such as sociological measures, political or occupational attitudes and family characteristics.

1. *Directory of Unpublished Experimental Measures,* Goldman, Bert A. and Saunders, John. Volume 1. New York: Behavioral Publications, 1974.

2. *Measures of Social Psychological Attitudes, Revised Edition,* Robinson, John P. and Shaver, Phillip R. Ann Arbor, MI: Institute for Social Research, University of Michigan, 1973.

3. *Measures of Occupational Attitudes and Characteristics (Appendix A to Measures of Political Attitudes),* Robinson, John P.; Athanasiou, Robert; and Head, Kendra B. Ann Arbor, MI: Institute for Social Research, University of Michigan, 1969.

4. *Measures of Political Attitudes,* Robinson, John P.; Rusk, Jerrold G.; and Head, Kendra B. Ann Arbor, MI: Institute for Social Research, University of Michigan, 1968.

5. *Scales for the Measurement of Attitudes,* Shaw, Marvin E. and Wright, Jack M. New York: McGraw-Hill Book Co., 1967.

6. *Family Measurement Techniques: Abstracts of Published Instruments,* Straus, Murray A. Minneapolis, MN: University of Minnesota Press, 1969.

Books

To locate books, begin with the card catalog at the library in which you are working; however, for information about other books, you will find the *Cumulative Book Index* (CBI) a helpful source. This is a list of most of the publications printed in English, excluding pamphlets and government documents. The CBI dates back to 1928. Other lists are available of books published prior to 1928, also. Ask your reference librarians for assistance; they are most helpful.

If your library doesn't have a particular publication that you need, ask if the *National Union Catalog,* or some other regional listing of books and periodicals cataloged by other libraries, is available. If it can be located through this search, you should be able to borrow the publication on interlibrary loan. Again, ask the librarians! They can provide a great deal of helpful information from other catalogs and directories through which they will likely be able to trace the particular volume you need.

Government Publications

Most of the publications of the U.S. Office of Education are listed in the *Education Index.* However, most libraries have the *Monthly Catalog of United States Government Publications,* which lists all publications of the various departments of the U.S. Government. Entries are made by department and also by subject.

Theses and Dissertations

The most complete source of dissertations completed in business education is the *Index to Doctoral Dissertations in Business Education,* 1900-1975, edited by Rahe (1975) and published by Delta Pi Epsilon. It lists 1,829 dissertations by author, by subject-matter classification, and by institution and year.

Theses and dissertations (both masters' and doctors') completed in business education are listed once a year in the *Business Education Forum,* published by the National Business Education Association. Prior to 1969 these lists appeared in each Spring issue of the *National Business Education Quarterly.* Also, once a year, these publications include abstracts of selected studies. However, the lists of business education studies are only as complete as the information submitted by the NABTE institutions; if no listings were submitted by a particular institution, nothing is likely to be reported from that school, so these are not necessarily complete listings and should not be relied on to furnish a complete bibliography of research on a topic.

Doctoral dissertations completed in many universities are listed in *Dissertation Abstracts* (Ann Arbor: Dissertation Abstracts, 1952-present). Prior to 1952 this directory was entitled *Microfilm Abstracts,* which dates back to 1938. Only the work done in participating universities is listed, but over the years the list of participating institutions has increased greatly. The roster of institutions having dissertations listed is given on the inside front cover of each volume.

Using the *Dissertation Abstracts* requires a bit of ingenuity on your part, particularly in selecting key words as they might appear in a title, since few business education subjects are listed. Therefore, dissertations in business education may be listed under curriculum, learning, business administration, methods, or vocational education, for instance, or under a number of other headings. It is, however, a very helpful publication, since an abstract of each dissertation is given in the volume. Thus you may get a good idea of whether or not the dissertation is likely to be one to include in your literature review.

The abstracts, however, tend to concentrate on the findings, and because of limitations of space, procedural details are usually not included. Thus these abstracts should never be used as **the** review of literature; the study itself must be read. All dissertations listed in *Dissertation Abstracts* are available on microfilm or in hard copy from Xerox University Microfilms, 300 North Zeeb Road, Ann Arbor, MI 48106. Beginning with Volume XXX, *Dissertation Abstracts* contains a Keyword Title Index in which the bibliographic entries are classified and arranged alphabetically by important keywords contained in the titles. A Cumulative Index is also now available for *Dissertation Abstracts,* covering the years from 1861 to 1972. One set of volumes lists the dissertations by author. Subject volumes are labeled by broad subject (education, business and economics, etc.) and then arranged by key word. Each issue of *Dissertation Abstracts* also has an author and subject index.

University Microfilms also operates a computer search service, DATRIX, which will provide a bibliography of dissertations on a particular topic. From a list of keywords that the researcher supplies, along with a brief descriptive summary of the goals of the proposed study, a bibliography of dissertations is made up by a computer search of titles. If none of the keywords is used in a title, that study may be overlooked; thus a little ingenuity on the part of the researcher will lead to a more complete listing.

Microfilms of dissertations in business education are available on interlibrary loan from the University of Tennessee Library. The listing of available dissertations was sent by Nu Chapter, Delta Pi Epsilon, to each of the NABTE institutions.

Phi Delta Kappa also publishes a series of references to doctoral dissertations, *Research Studies in Education* (Mary L. Lyda, *et al.*, eds., Bloomington, IN: Phi Delta Kappa) 1953-present.

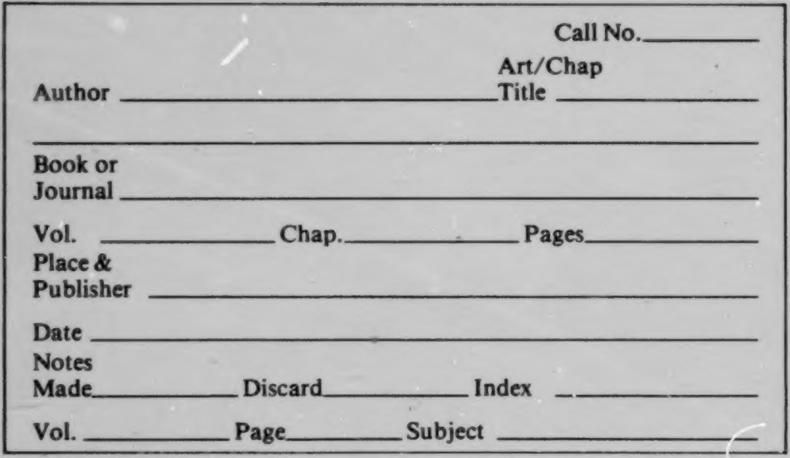
Other research projects, not dissertations, are listed in ARM and AIM, publications of the ERIC system. Copies of these materials are available on microfiche from ERIC, Vocational and Technical Education Division.

Preparing and Using a Master Bibliography

As you search the indexes and catalogs for likely materials to be read, certain procedures will save you a great deal of time and will help to make the reading process more efficient.

Prepare a separate slip for each title you think merits your consideration. Compiling your bibliography on cards, not on sheets of paper or in a notebook, is essential. It will save you much time and wasted effort later on.

Duplicate a standard form for recording the bibliographic data. Using a form such as illustrated in Figure 2-2 has many advantages. It will help you to record **all** the information necessary for the final bibliography or list of references that you will include in your report. Recording all information at



Call No.__________

Author ____________________ Art/Chap Title ____________

Book or Journal ____________________________

Vol. __________ Chap.__________ Pages__________

Place & Publisher ____________________________

Date ____________________________

Notes Made__________ Discard__________ Index __________

Vol. __________ Page__________ Subject __________

Figure 2-2. Master Bibliography Card

this time will save you precious time later. You will not be as likely to have forgotten the page number, volume number, or some other piece of information if you fill in a standard form. Forgetting one of the items means another trip to the library; and if you have not recorded the particular index or catalog from which you obtained the listing, you can spend a great deal of time tracing it.

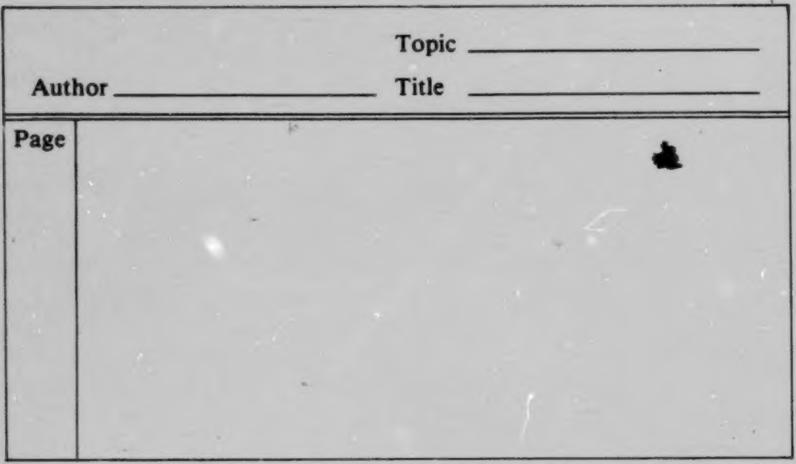
Make two copies of your master list. Cut carbon paper to fit the bibliography slips and make two copies of the basic information, keeping one set in alphabetic order by author and the other in "reading" order. As you finish filling out each slip, file one set in alphabetic order by author and the other set according to the book or journal in which the article or chapter appears. File the cards for each journal in chronological order, with the most current references first. This systematizes your reading, since you can concentrate on one section of the library at a time, reading all the articles from one journal before proceeding to another. This is a real timesaver. Then, by having the other set of cards in alphabetic order by author, you can check each bibliographic entry against your alphabetic list as you go through each index. You will save yourself from making a second notation and looking up an article a second time. You are likely to come across some articles in several diferent indexes, but you don't always remember that you have seen it before; thus checking your alphabetic file saves you time.

Reading the Literature and Making Notes

To proceed with the actual reading, use the cards arranged by journal; you will save yourself much time by developing efficient reading techniques. Scan the article first and then read the summary. Scanning and reading summaries allows you to evaluate the usefulness of the article. Use a similar procedure in reading dissertations and other research reports. Scanning the first chapter and reading the last one will give you enough information about the research to know whether it merits further consideration. At this point you can eliminate the item if it will make no contribution to your study. Mark that fact on the author card and **keep the card in the author file.** If you fail to keep the slip you have marked "Discard," you will find yourself later rereading something that you had at one time decided was of no value. If you read an item and make notes, mark the author card "Noted" or "Notes made." This way you have a record of exactly what you have read, and again you will save duplication of effort.

Make additional bibliography cards as you find other promising titles in the bibliographies of the materials you read, checking each one with your author file, and inserting the copies in the appropriate places in your two decks of cards.

All notes should be made on **cards.** They are easier to handle and easier to arrange in the order that will be of most help to you in writing the report. Most people find either 4 x 6-inch or 5 x 8-inch cards the most suitable size for making notes because they are large enough for quite extensive notes and yet are small enough to handle easily. Lightweight cards are much easier to handle than paper is. Duplicating a standard note card will make reading and notemaking and your later writing more efficient. Figure 2-3 illustrates such a form.


Topic ____________________

Author ____________________ Title ____________________

Page

Figure 2-3. Reading Notes Card

Complete bibliographic information is not required on the note card; just record the author's name, the journal (book), and date, and enough of the article or chapter title to allow you to identify it in your master file. In the upper right corner record the particular topic with which that part of the work deals, **using a separate card for each topic** discussed in the piece you are reading. This makes possible easy organization of your note cards according to topic, in readiness for writing the literature review in your report. For instance, if you are reading a report of an experiment for which the psychology of learning was the theoretical base, you might have one card concerning distribution of practice, another on the role of feedback, and a third one regarding incentives, even though each card contains only four or five lines of notes.

A most important step in making notes is to **record in the left margin the number of the page** from which you are taking notes. Then each time you turn the page as you read, record the next page number!! This will save you extra trips to the library and much search time.

How extensive should your notes be, or should you copy verbatim? Paraphrasing has the advantage that it makes you think about what you are reading. You must understand what is being read in order to make a good paraphrase, and you save time not having to copy quotations. However, if some statement made by the author is so succinctly and uniquely stated that you could not keep the author's meaning in a paraphrase, then a quotation may be called for. A literature review that is but a series of quotations connected with a phrase or sentence or two may be interpreted by some readers as either a sign of incompetence in understanding what was read or a sign of intellectual indolence.

As you take notes, record your own comments and/or questions about what you are reading, preferably in another color ink. This will help you later in evaluating the piece you are reading, and you will not lose some of the thoughts you may have had about some important point. As you read, always think about what you are reading in relation to your particular research problem regarding procedures or logic, and record your thoughts about them. However, do make complete notes (in outline form) regarding procedures used in research studies, since findings of research must always be considered in view of the procedures used and the underlying theoretical base. Detail is important in evaluating procedures.

Systematizing and organizing the preparation of the bibliography and the reading and notemaking will save you much valuable time and effort and will save you from many frustrations that are common among researchers in searching, reading, and reporting literature.

Chapter 3

Data Gathering

During the process of delineating the problem, the researcher must decide on the specific types and amounts of data to gather, as well as where and how to collect them. Are tests, scales, or inventories available that will serve your purpose, or will you have to develop your own instruments? This step requires careful, detailed consideration.

Planning the Data Gathering Activities

One helpful technique for systematizing your planning is to determine for each objective and/or subproblem the exact data needed by making a worksheet similar to the one in Figure 3-1. It contains several columns that may be filled in as your thinking and reading progress. For each of the research questions or subproblems, record in the first column all possible types of data you may need for answering the particular question. The second column is the place to record the contribution the data may make in solving the problem. State why, specifically, you need that particular set of data (in other words, justify the data). In the third column, list the methods and sources for collecting the data. Will it be collected by questionnaire, published instrument, interview, observation, etc.? Also indicate whether the data will come from students, teachers, parents, alumni, or some other source. Perhaps they will come from school records, public records, or from reference books.

Selecting Data-Gathering Instruments

Certain considerations and steps must be taken into account in selecting and defining variables and the types of data to be collected:

1. **Operationally define each set of data.** In looking at the chart in Figure 3-1, we see that the types of data are stated only in general terms. Each one must be operationally defined. For instance, exactly what do you mean by general ability level? General ability level as measured by what instrument: *California Test of Mental Maturity? Otis Quick-Scoring Intelligence Test? Scholastic Aptitude Test?* Which readability formula will define the readability level of the materials most satisfactorily? What type of reading ability? comprehension? speed? a combination score? as measured by which test? As you will recall from Chapter 1, data and objectives (subproblems) are interdependent. Objectives determine the data you need, and the availability of data governs the subproblems that can be included in the study.

No data should be gathered that are not called for by the problem or any of the subproblems. Sometimes one is tempted to get one more piece of

Variables Data	Why	Sources
Sex	Boys may react differently from girls to textbooks and newspapers	Teacher's records
Ability level	May be related to achievement in money management unit	Standardized tests—what scores are in school files? Nelson-Henmon Calif. Test of Mental Maturity
Reading ability	May influence amount gained from reading assignments	Iowa Achievement Tests—Reading See Buros Mental Management Yearbooks
Grade level	Effect of maturity; immediacy of need for topic	School records Teacher
Economics background	May have effect on learning re money management	Questionnaire to students? School records
Other classes in which topic may be covered	Background could account for knowledge instead of reading assignments	Questionnaire?
Reading materials	Interest Information	Textbooks Magazines Newspapers
Achievement in money management unit	Criterion	Make up one Publisher Other studies
Attitude toward class	Influence on interest?	Questionnaire? Published instrument? See Buros
Motivation	Influence on achievement	Published instrument See Buros
SES	Attitudes and knowledge of various socioeconomic status levels may differ and have effect	Standardizing SES scales Students?

Figure 3-1. Planning Sheet for Determining Data to be Collected

information "because it would be so nice to have" (perhaps for another study). Resist that temptation.

2. **Establish the validity of the data-gathering instrument(s).** All data collected must be **directly** related to the problem. Validity is perhaps the most important characteristic of the data. Data must provide exactly the information that you are seeking. In the case of published tests and scales (attitudes, interests, personality, etc.), validity figures usually are available for the particular instrument. However, this is not a guarantee that the instrument is valid for the particular group you are working with in that particular situation. If validity of an instrument has been reported for groups of post-secondary students, it may have little or no validity for a group of ninth or tenth grade students.

To measure what the students learn from a unit on money management in our illustrative study, the instrument must measure only those aspects of personal finance delineated by the objectives for that part of the course. If we use a standardized test or an instrument prepared by another researcher for a similar study, we must be sure that it covers the same materials and objectives as those under study in our research. This is known as **content** validity. This is often established by having a panel of "experts" evaluate the coverage of the instrument in relation to the objectives set forth for the unit or course.

Some other types of validity may be more appropriate in other situations. Perhaps **concurrent** validity is appropriate in one situation, while **predictive** validity is the choice for other purposes. To establish concurrent validity, the scores on one instrument are correlated with the scores by the same individuals on a similar, well-established instrument. The two instruments are administered at approximately the same time. On the other hand, predictive validity is the relationship between the instrument and some future criterion measure; in other words, how well does it predict the subject's performance on some other measure at another time. For instance, if we give students a five-minute straight-copy typing test today, we can predict quite accurately what the student's rate is likely to be tomorrow on a five-minute test of similar difficulty. A five-minute straight-copy test has predictive validity—for determining what a student's score is likely to be on another five-minute straight-copy test. However, this degree of validity does not hold if we use the straight-copy typing test to predict scores on a 20-minute business-letter production test or for predicting transcription rates or skill in setting up tables. No test or other instrument is valid in and of itself—it is valid only for particular purposes under specified conditions.

Yet another type of validity is appropriate for instruments designed to measure hypothetical (theoretical) constructs; that is, such things as anxiety, motivation, self-image, or attitudes. An oversimplified, limited definition of **construct** validity is that the instrument does not measure something other than what it is intended to measure. Perhaps an illustration will help clarify this idea. Beattie (1962), in his study of the relationship between attitudes of students and their knowledge of four phases of personal finance, developed an instrument to measure attitudes and another to measure knowledges. He wanted to be sure that terminology was not a barrier to the students' understanding of the items on the attitude scale—he did not want lack of knowledge or low reading level to interfere with measuring attitude. He did not want to measure reading ability with this attitude instrument. Thus, in try-outs of the instrument, he had students mark all words they were not absolutely sure they understood. He then revised the items to simplify the vocabulary. He also applied a readability formula deliberately structuring the scale for no higher than the sixth grade, even though the instrument was to be administered to ninth and tenth graders.

Another interpretation of construct validity is provided by Borg and Gall (1971): "Construct validity is the extent to which a particular test can be shown to measure hypothetical constructs" (p. 141). They illustrate this by explaining the development of an instrument designed to measure anxiety. It should correlate highly with physical manifestations of anxiety in individuals in stress situations; e.g. higher than normal blood pressure, perspiring palms, rapid pulse, etc.

Since no data-gathering instrument is valid in and of itself, but is valid only in given situations, its validity can vary from place to place, from group to group. Thus, an essential step in selecting or developing a test or other data-gathering instrument is to establish its validity for the particular purpose for which it is to be used in one's own study. No test should be used in a study unless its validity for the particular research situation is demonstrated and reported.

3. **Determine the reliability of the data-gathering instrument(s).** A second important quality of any data-gathering instrument used in research is its reliability; i.e. its stability or consistency of measurement. As with validity, reliability may be determined in several different ways, depending on the nature of the measuring device.

For untimed instruments that result in a total score, a **split-half** technique may be appropriate if the items are independent of each other; that is, if the answer to one item does not depend upon having answered a previous item. Essentially one compares (correlates) the scores students make on half the items with the scores they make on the other half of the test. The test is split in half by comparing the scores on the odd-numbered items with the scores made on the even-numbered items. Since this technique depends on **all** the items in the test, and since most students do not complete a timed test, this is not an appropriate method of determining reliability for timed tests. It may not be appropriate for determining the reliability of questionnaires since the items are neither correct nor incorrect and many such instruments do not yield a total score.

In such situations a **test-retest** procedure may be more suitable. An instrument may be administered today. At another time the same instrument is given again to the same subjects, and the results of the two administrations are compared. Did the subjects get approximately the same score on both administrations; that is, did they answer questions on a questionnaire in the same way both times? The time lapse between administrations of the instrument depends on the situation and must be decided upon according to one's best judgment. If memory might be a factor influencing the responses on a second administration of the instrument, one should wait until that factor is likely not to have too much effect. On the other hand, the time between administrations (especially for achievement tests) should not be so great that additional learning or skill development that has taken place in the interval between administrations may change the scores to any great extent. In the case of a questionnaire regarding attitudes, the danger exists, of course, that something could occur between administrations that would drastically affect the subjects' attitudes toward the topic. On the other hand, if memory is not an important factor, as in straight-copy typing tests, they can be administered immediately one after the other.

Correlating **alternate forms** of the instrument, applying one of the Kuder-Richardson formulas or by using an analysis of variance technique are other ways to establish reliability, but these are beyond the scope of this bulletin. They are discussed in several educational measurement textbooks (see the List of References at the end of this Bulletin) and are important and widely-used methods for determining reliability.

One point for novice researchers to keep in mind is that in most measurement situations one should be concerned with validity to a greater degree than reliability, important as that is. Many measures can be extremely accurate (reliable) but not be valid. The following illustrates this point:

The students in a class wonder what the mean weight of the class is; no scale is available. However, an accurate steel ruler is found in the room. Someone suggests that tall people generally weigh more than short people, and they decide to measure each person's height. They take the measures twice, and find them to be very accurate; they correlate almost perfectly—thus their measures are highly reliable. The class then estimates each person's weight from the individual's height. Someone brings a reliable scale to the next class meeting. Upon weighing each individual, the class members find their estimates in error by a considerable margin. Measures of individuals' heights, reliable as they may be, do not give a very accurate measure of their weights—it is not a valid measure of weight. Recall also the high reliability of straight-copy timings in typing, but their low validity for measuring letter production or the speed with which a student may be able to type tables.

On the other hand, the validity of a measure is no higher than its reliability. Thus if we have a highly valid measure, we can use it with some assurance that it is reliable.

4. **Consider the ethics of data gathering.** Heimerl and Halldorson (1971, p. 304) have listed five steps to take in order to assure ethical data gathering in research situations:

a. Tell the subjects the purpose of the research and the use to be made of the data.
b. Assure the subjects that no unpleasant effects will result from the procedures used in the study.
c. Assure the subjects that all data gathered are confidential and will not be in any way identified with any individual handling the data.
d. Never ask the subjects to perform in an unprofessional manner.
e. Assure the subjects that they may receive a summary of the analyzed data.

In most school systems anyone wishing to gather data must get approval of the research project and permission from the administration to collect data before going into the classrooms or to the teachers for information.

The purpose of the study and the use to be made of it must be explained and copies of the data-gathering instruments should be shown to the administration.

Recent legislation regarding human rights is also affecting the availability of certain kinds of data. Not only must permission be given by the school administration, but the persons under study must give their permission for certain private and personal data, such as grade-point averages and other information from students' files. Universities over the country are establishing human subjects research committees to evaluate all research proposals, including any data-gathering instruments, for all studies involving human subjects. This could have a profound effect on the kinds of research that may be conducted in the future.

Advantages and Disadvantages of Survey Data-Gathering Techniques

The following sections outline advantages and disadvantages of using the various data-gathering techniques and instruments in survey research. Some procedures will be outlined that may reduce the problems and increase the utility of survey (status) studies.

Well-done survey studies serve several purposes. They are good exploratory devices to find out what the current situation is—the state of the art, so to speak, from which hypotheses may be developed for other more detailed, definitive studies. If the researcher is interested in a new area or new procedure about which very little is known, a survey may be helpful in discovering precisely what needs to be studied in depth by some other means. It may provide clues to possible procedures that need to be tried out before curricular or methodology decisions for the classroom are made through experimentation. Although surveys may reveal opinions regarding areas of educational situations that need improving, such opinions do not provide adequate bases for **how** the improvements should be made (Erickson and Oliverio, 1964, p. 5; Borg and Gall, 1971, p. 192; Galfo and Miller, 1970, p. 26). As Mouly (1970) points out, opinions and attitudes, and even collections of facts, from a large group of respondents still do not establish cause and effect:

> Frequently, questionnaire research constitutes nothing more than a pooling of ignorance; the opinion of one single expert may be far superior to the compilation of the opinions of many persons who do not have the information and/or who are incapable or unwilling to convey information they have. (p. 261)

Data from surveys, however, may show relationships that can be studied further by other means that can provide information about cause and effect that may provide the basis for policy decisions.

Any survey must be more than just a data-gathering exercise. The fact that many surveys are rather superficial sets of questions about superficial topics has been a common criticism of such studies. The problem one selects for a survey must fulfill the criteria of a good problem listed in Chapter 1: it must ask a problematical question; it must include more than one variable; it must not contain value statements; and the data must be collectible. Yet, even the question that involves two sets of variables and compares them may not really be a problem.

Consider a proposal of a study to discover whether office procedures teachers were up to date in their knowledge of typing practices required of beginning typists in offices. The data were to be analyzed according to whether or not the teachers had advanced degrees, the amount of work experience and the amount of teaching experience of the teachers, the number and kinds of workshops attended, and other professional experiences of the teachers. The question was concerned with the relationships among variables; but was it really a problem? Was it in reality a data-gathering exercise? What difference does it make if teachers are aware of how things are typed in an office (for instance, that most tables are set up by judgment placement, according to Ober's (1974) study), if these teachers instruct their students always to calculate the placement or to use the backspace-centering method to set up tables exactly and never allow students to set columns by judgment? The study was strengthened a great deal, in spite of the good questionnaire that was designed, by asking students the same questions regarding typing practices that were being asked of the teachers so the responses of the students could be compared with teachers' responses. One must be sure that even a complex survey serves a greater purpose than merely collecting data, unless one is exploring a new field about which very little, if any, information is available upon which to form the bases of other studies.

Another common criticism of questionnaire surveys is that they do not allow in-depth examination of the questions. Unless quite sophisticated instruments and techniques are used, relatively superficial or "socially accepted" attitudes tend to emerge. Use of in-depth techniques requires considerable background knowledge and sophistication in developing measuring instruments and in theoretical constructs of personality. However, certain techniques have been developed that may be more suitable than direct questions for eliciting attitudes and feelings; these will be discussed later in this chapter.

Galfo and Miller (1970, p. 26) point to two common fallacies in the use of survey data:

1. A common fallacious belief is that a widespread practice is "good," or "what ought to be." A survey of a group of typing teachers regarding straight-copy standards may reveal that x wpm is the mean rate for a C grade. However, several questions come to mind when such results are reported: How representative are the teachers? How good are they at teaching typing? What methods for speed building do they use? If they had used other speed-building techniques, would the mean rates be different from those reported? What was reported (even if the sample was of 10,000 teachers) may not be the standard that "should" or could be. Perhaps had

the teachers used other teaching methods, their students may have attained higher rates than those reported.

2. People tend to believe that the truth of a "fact" is directly proportional to the number of people who accept its accuracy. Many readers of research seem to feel that if a great enough group of people say something, it must be true. This brings to mind Columbus, who, if he had followed the opinions of the populace, perhaps would never have attempted the voyage that brought him to America. The lone thinker just may be right; and only through experimental research may the value of new approaches or practices, materials or methods be ascertained.

With this amount of negative criticism of surveys having been called to your attention, one may wonder whether such research has any value at all. For certain purposes, it is the research method of choice. It is a valuable exploratory tool for finding out what people know, what they like and dislike, what they think (attitudes and beliefs), what has taken place, and what is happening now, the status quo. To gather data about these matters, a variety of data-gathering techniques and instruments are available. Some of these things are best found out by questionnaire, others by interview, and still others by already prepared instruments designed for the purpose. So let us look at each of the data-gathering techniques.

Questionnaires

Besides gathering information by means of paper and pencil tests and scales, data may be collected in a variety of other ways. Perhaps the most widely used instrument is the questionnaire in which straightforward questions are asked of the respondent. Other instruments and techniques, such as a semantic differential or a Q-sort, may be more suitable than a questionnaire where the researcher feels that a more subtle approach may produce more valid, reliable responses. Other studies may more appropriately call for a task analysis or the use of a Delphi technique. Studies of classroom behavior may call for actual observation of teacher and student behavior.

The ubiquitous questionnaire has been used and misused perhaps more than any other data-gathering technique. It is the instrument of choice when a large sample is required and one's finances are limited. It is the instrument of choice, however, only after other techniques have been considered and rejected for good reasons.

General Considerations in Using Questionnaires

Questionnaires are a convenient way to get factual data, provided they cannot be obtained any other way. One should never ask for information that is available from other sources. For instance, rather than asking school administrators in one's own district for information regarding tax rates, consult city and/or county records where these data are easily available. Rather than asking leaders in business education for a list of their writings, consult *Business Education Index, Education Index, Business Periodicals*

Index, or other appropriate references. Biographical data of educational leaders are often available in one of the *Who's Who* volumes. The Census Bureau will supply, at cost, data about population characteristics of many geographical divisions, including cities and counties.

By using mailed questionnaires, one usually can collect data from larger, more representative samples than one can usually get by interview. If questionnaires are properly designed, though, more candid and objective replies may be obtained than by other means. However, to a certain extent this may depend on the skill of the interviewer or the design of the questionnaire and the topic under consideration. The questionnaire does have the limitation, however, that one cannot evaluate whether the respondent is really being evasive, is uncomfortable while answering the questions, or reluctant to give answers. One cannot help a respondent understand the questions as one can in an interview. (Interviewers must be trained, however, in helping respondents, not to "lead" them to the type of response the interviewer believes should be given.) On the other hand, some persons claim that the fact that the questions are uniform, that everyone has exactly the same questions given in the same way, is a strength of the mailed questionnaire.

Questionnaires may allow the respondent more time to think about his answers than he has in an interview. This depends to some extent, though, on the willingness of the respondents and their ability to answer the questions. If the subjects feel that the topic under study is foolish or about some inconsequential matter, they either may not answer the questions at all or they may just mark anything, giving little thought to their responses. Other respondents may try to "help" the researcher by answering the questions the way they think the researcher wants them answered. In other circumstances, respondents may answer the questions even though they may not really know the answers (they do not want to appear ignorant) or they may not recall accurately the answer or will not take the time to look up necessary data. In one survey of reading habits of people in New York, nearly 10 percent of the people said they currently read *Collier's* regularly even though it had ceased publication seven years before the survey was taken.

The most serious criticism of the use of questionnaires, however, is the bias in the results due to nonresponse. This problem is one that is neglected by most novice researchers. It is seldom checked; at least it is seldom reported. Evidence indicates that often nonrespondents are different from those who do return completed questionnaires. This is particularly true in follow-up studies of graduates, especially when some questions involve socio-economic status. In some cases, late respondents also are different from those who responded soon after receiving the questionnaire. However, an important step in using questionnaires is to compare the respondents with the original sample to whom the questionnaire was sent. For instance, one usually knows something about the characteristics of those who constitute the sample to whom the questionnaire is sent. If one is plannning to compare attitudes of business teachers toward new teaching techniques and is planning to compare teachers from large-, medium-, and small-sized high schools, one knows how many teachers in the original sample teach in schools of each size. One MUST compare the proportion of respondents from large schools, with the proportion of the entire sample from large schools; and similar comparisons must be made for teachers in other sizes of schools also. If the proportions are about the same (statistical tests should be made to ascertain any lack of similarity), then one has some idea of how representative of the entire sample the respondents are. If they differ, the findings must be interpreted in light of that information, and perhaps no definite conclusions should be drawn from data from such a sample. Unless the response rate is over 80 percent, sample bias is likely to be quite pronounced.

The researcher can take several steps to increase the response rate: develop a good questionnaire; select an appropriate sample; write an appealing cover letter; and pursue nonrespondents.

Developing a Good Questionnaire

Once you have selected your problem and decided what information you need, you are ready to start formulating questions and putting the questionnaire together. Several steps are given here that will help you develop a valid, reliable questionnaire that encourages responses. The steps are not necessarily in chronological order, since at times two or more steps may be in progress simultaneously.

1. **Visualize the respondents.** In drafting the questions, try to visualize your respondents, how they may react toward the study, how they might interpret what you have written. Try to think of every possible way they might answer the questions. Have your friends read into the questions everything they can and try to misinterpret them as much as possible and answer them accordingly. Some respondents somewhere will interpret questions in ways you never thought possible, so make your questions as clear and specific as possible.

2. **Group together questions dealing with each aspect of the study.** List for each subproblem all the questions you need to ask in order to get the necessary data. Ask other people to help you think of things to ask. Sometimes a person who is not too close to the situation, or someone from another field entirely, will ask questions you will not think of; they ask questions because they don't know the answer. These often are valuable questions.

Examine the questions to see if they might be grouped more logically according to types of information rather than according to subproblems. Subproblems of a study may deal with reactions to situations by men and women teachers; teachers with different amounts of experience; teachers with a bachelor's or an advanced degree. The responses sought, however, may be to questions regarding problems of classroom management, students' personal problems, maintaining discipline, administrative support, motivating students, etc. Questions dealing with teachrs' personal data should be placed together, and those questions dealing with each of the

Competencies Needed by the Teacher of Shorthand

DIRECTIONS: Please indicate your opinion of the relative importance of each item as a competence of the teacher of shorthand.

C = Crucial
H = Highly desirable
D = Desirable but not absolutely necessary
N = Not necessary
I = Incorrectly stated, needs revision
X = Do not use; concept inappropriate

Planning:

The shorthand teacher will:

1. determine how the shorthand course fits into the total school program and its relationship to other courses. 1. CHDNIX
2. determine how the shorthand course fits into the business education program and its relationship to other business courses. 2. CHDNIX

.

.

.

38. develop an effective plan for a typical day of "dictation-transcription." 38. CHDNIX

Content:

The shorthand teacher will:

1. know brief forms and derivatives. 1. CHDNIX
2. know theory (word endings, beginnings, blends). 2. CHDNIX

.

.

.

20. select supplementary materials. 20. CHDNIX

Administration:

The shorthand teacher will:

1. demonstrate the ability to conduct conferences (parents, teachers, students). 1. CHDNIX
2. effectively handle discipline problems. 2. CHDNIX

.

.

.

19. develop criterion based levels of competencies. 19. CHDNIX

Illustration 3-1. Grouping Questions

problem categories should be grouped together and each group labeled with an appropriate heading as in Illustration 3-1. The questions in Illustration 3-1 come from a rather long questionnaire asking for opinions of teacher educators regarding competencies needed by shorthand teachers. The questions on this 12-page questionnaire were grouped into eight areas, of which three are shown in the illustration. Organizing the questionnaire this way helps the respondents "shift gears" in thinking about the problem situations.

3. **Arrange questions in either a psychological or logical order.** The questions should flow from one to the other smoothly. First establish rapport by asking easily answered, neutral, interesting questions. If the questionnaire is quite long, put important questions nearer the beginning than the end. Respondents may get tired or bored with answering questions and not give the last questions as much thought and care as they did the first ones.

Questions should be arranged from general to specific. If specific questions are asked first, then the general question is likely to be answered only in terms of the specific answers, which may not be what you really are looking for. Notice that Question A in Illustration 3-2 asks for an over-all evaluation of the respondents' summer school experiences, while Questions B and C call for specific things about the summer session that impressed them one way or the other. Had the last two questions been placed first, the answers to Question A may have been influenced by the responses to the first two questions.

IV. GENERAL APPRAISAL

A. Considering your chief reason for attending Summer Session this year, how would you evaluate the experience?

Very well satisfied; quite certain that I got what I wanted ______1

Fairly well satisfied; might have been better in some respects, but I am glad I attended ______2

Uncertain of the value of attendance this summer for me ______3

Somewhat dissatisfied; perhaps I should have gone elsewhere or done something else ______4

Very dissatisfied; I shouldn't have attended this summer session ______5 (42)

B. Was there any phase of the Summer Session program that struck you as being particularly outstanding? If so, please describe it:

C. Was there any phase of the Summer Session program that you found to be particularly unsatisfactory? If so, please describe it:

Illustration 3-2. Arrangement of Questions from General to Specific.

Sometimes a general question can be used as a classification item as in Illustration 3-3. If it is answered one way rather than another, the respondent may skip the next few questions, thus saving both time and effort for the person answering the questionnaire. It also helps you, the researcher, in processing the data, since the questionnaires can be sorted according to the response to the classification question. This saves your having to handle all the questionnaires for that set of questions. Chances are that a smaller number of summer school students attended the university during the regular academic year than did not. Thus, all the data cards for those not attending school during the academic year may be eliminated, requiring the handling of only a portion of all the cards for these three questions.

A. Did you attend the University X during the 19— - 19— academic year?

Yes________

No________

B. If your answer to question A is yes, please answer questions 1, 2, and 3 below; otherwise skip to section VI on page 5.

1. In general, how do the work requirements of Summer Session courses compare with those of courses that you have taken during the regular academic year?

Courses are more demanding in the summer ______(1)

Work expected is about the same for courses in the regular academic year ______(2)

Courses are less demanding in the summer ______(3) (46)

2. In general, how does the quality of instruction in Summer Session courses compare with that of courses that you have taken during the regular academic year?

Generally better in the summer ______(1)

About the same ______(2)

Generally poorer in the summer ______(3) (47)

3. Do you think that you learned as much in the intensive Summer Session courses as you would if you had taken the same courses during the regular academic year?

Learned more in the summer ______(1)

Probably about the same ______(2)

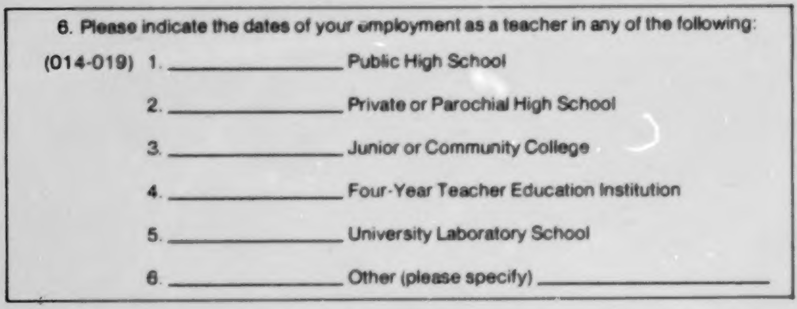
Learned less in the summer ______(3) (48)

Illustration 3-3. Classification Questions Reduce the Number of Data Cards Handled.

4. **Make apparent that the questions are related to the purpose of the study.** Do not ask for some information "because it would be good to have it." Unless it is directly related to the problem, resist the temptation to include a question. Also ask no questions that will seem foolish or unimportant to the respondent. Ask the questions in such a way that their purpose in relation to the problem is evident. If you need to ask a question that may seem "far-fetched," try to "warn" the respondent.

5. **Use an easy-to-answer format.** By asking questions that are understandable and easy to answer, the number of responses may be increased. This may be done in several ways:

a. **Decide on the form the questions should take.** Decide which questions should be open-ended and which ones will allow appropriate check-off responses. Open-end questions may give the respondents more freedom to express themselves, but many people may not make the effort to write out an answer, or they may respond only superficially. Also, categorizing the responses to open-end questions may be quite difficult; they may not fit into categories you may have anticipated. On the other hand, presenting a list of possible responses may restrict respondents from giving answers they might have given had they not been "directed" in their thinking by the list of possible responses. This fault may be partially overcome by including a choice of "Other (please specify________)," as in this question:



6. Please indicate the dates of your employment as a teacher in any of the following:

(014-019) 1. ________________ Public High School

2. ________________ Private or Parochial High School

3. ________________ Junior or Community College

4. ________________ Four-Year Teacher Education Institution

5. ________________ University Laboratory School

6. ________________ Other (please specify) ____________________

Illustration 3-4. Providing a Space for Unlisted Responses.

No provision is made in the responses for business teachers who may have taught in a vocational high school or in a proprietary business school, for instance. Providing the "Other (please specify________)" choice allows those with such teaching experience the opportunity to respond.

b. **Use the simplest language possible.** Avoid technical terms; not all respondents may be familiar with them, or they may have different interpretations of the terms. Defining terms or specifying exactly what you mean may help all the respondents to interpret your questions in the same way:

> In your opinion, what is the impact of the SBA loan program on your community (**Impact** here is defined as preserving free enterprise and strengthening the overall economy by aid to small business)?

Consider the term "secondary education," which may refer either to grades 9 to 12 or grades 10 to 12. If you ask a question like "What is your salary?" respondents may have any of a number of questions about it. Does that mean annually, monthly, weekly, or hourly? Does it include extra pay for extra-curricular duties and/or evening school classes? Suppose you ask the simple question, "What is your age?" Do you mean at the nearest birthday, last birthday, or next birthday? How exact do you want the information, to the nearest year or month?

Also, keep the vocabulary free from jargon, using the simplest word you can. Use **teaching** instead of **pedagogical**, for instance. Or what is wrong with using "yes" or "no" to indicate whether or not one is willing to cooperate in the following item:

> If you checked either response 1 or 2 above, please indicate your willingness to share a copy of your work to assist in the development of the plans of the University of______________.
>
> __________ Affirmative   __________ Negative

c. **Use precise wording that asks exactly what you want to know.** A census questionnaire contained a blank labeled, "Length of Residence." One reply was "50 feet," when the information wanted was the length of time the individual had lived in the community.

In an instrument asking high school office occupations students to solve some typical problems that occur in an office, the question, "In this situation, what would you do?" is more descriptive of the kind of response you want than "How would you respond?" Some students might think of "respond" only in terms of oral responses rather than in terms of actions to be taken.

Specify age in a question like this:

> In which of the following age brackets were you as of July 1, 1976?
>
> Under 20 years ________
> 20-24 years ________
> 25-29 years ________
> 30-34 years ________
> 35-39 years ________
> 40-49 years ________
> 50-59 years ________
> 60 years or over ________

d. **Use no "loaded" or "leading" questions.** A question should not indicate to the respondent an answer he or she should give. Asking students "Why don't you like the consumer education class?" may imply that generally students dislike at least something about the class. In fact, they may rather like it, but because of the way the question is worded, students' responses may not reflect their true opinions about the class.

e. **Questions must not be ambiguous.** Some questions seem to have different meanings depending upon how they are read. Suppose you ask, "Do you make announcements at the beginning or end of the period?" Some readers may think you are asking whether they make announcements, while others may think you are asking at what time in the period they make announcements.

f. **Questions must be answerable.** Consider this question: "Do you feel that office procedures should be taught in 3- or 4-hour blocks, or should they be taught in shorter blocks, yes or no?" Obviously that question cannot be answered either "yes" or "no." This is a situation where some sort of check-off multiple choice response is appropriate and practical. The question may take the form:

What time arrangement do you feel is most practical for teaching office procedures?

3-4 hour block, 3 times a week ________
3-4 hour block, daily ________
2-hour block, 3 times a week ________
2-hour block, twice a week ________
Single period, daily ________
________________________ Other, please specify ________

Consider the following question:

Do you introduce writing in shorthand by having students scribble-write, dry-pen, or trace? Yes________ No________

How would you answer that question? Does it mean do you use all three, more than one, or only one of the three techniques? What if you used some other technique? Would the question be easier to answer if it read:

Do you introduce writing in shorthand by having students use any of these techniques: Scribble-write, dry-pen, or trace? Yes________ No________

Would you rather answer the question in this form?

Which of the following techniques do you use in introducing writing to your students in beginning shorthand:

Scribble-writing . . . . ________

Tracing . . . . ________

Dry-penning . . . . ________

______________________________ Other, please specify . . . . ________

g. **Use check-off items whenever you can.** Check-off items in which the respondents mark their answers in boxes or blanks are very effective when you can supply a complete list of alternatives. Free response questions may require too much work to complete. When the choices are relatively limited, the better procedure is to ask the respondent to make a check mark, to circle a letter, or to cross out or underline than to ask for a written response. But be sure that all the most likely responses are included and include an "other, please specify" option for those who do not feel their answer fits any of the listed categories.

A researcher, trying to determine the interests and primary social activities of two groups of secretaries, asked a question about the spectator sports events they attended or watched on TV. The study was done in a state where football is of major interest—but through oversight it had been omitted from the list of choices. However, an "Other, please specify" space had been provided, and "football" was written in by most of the respondents.

If you ask the respondents to rank a list of responses, the list should probably contain no more than seven items. Few if any people can clearly distinguish more than seven gradations of anything.

The list of choices from which the respondent selects a reply should be in the same mode, i.e., all verb form statements, all beginning with a noun, personal pronoun, or adjective.

To check whether or not you have an adequate list of choices to which individuals can respond, you could ask this question in open-end form to a smaller group of persons similar to the population from which you are going to sample. Check how well their responses fit into the response categories you have established, and if necessary revise the response list you planned.

h. **Have responses that relate directly to the question.** Of course, all responses listed must be directly related to and answer the question as stated. The question below appeared on a questionnaire entitled "Shorthand Issues," for which the instructions explained that the issues were those that affect administrative concerns and that methods in teaching the subject matter were excluded. The instructions also said to limit the response to one best alternative. However, checking more than one alternative was allowed if absolutely necessary to reflect the respondent's opinion.

The shorthand listening laboratory has had what effect on shorthand instruction?

________ (a) none on improving student achievement

________ (b) a negative effect when used as the total method

________ (c) provided rest and relaxation periods for teachers

________ (d) allowed for better treatment of individual differences

________ (e) been well suited to providing dictation for homework

________ (f) other, please specify ________________

Is Response (c) really concerned with effect on shorthand instruction? Does it reflect a bias on the part of the investigator? What about Response (b)? How many teachers use the shorthand lab as the total method? How many readers are likely to catch the "total method" portion of that statement? Are Responses (a) and (c) really the same? Notice that all negative effects are listed first. Since this is the last question on a six-page questionnaire, the respondent may not read this question very carefully, and the response is likely to be "loaded" negatively as a result. No provision was made for effects of the laboratory if it is used in combination with other instructional methods. Reframing the question, giving it a more neutral tone, could draw more responses and give a more accurate indication of the primary effect of a shorthand laboratory, as in this question:

What has been the most important effect of using the shorthand listening laboratory in shorthand instruction?

________ (a) had no effect on improving student achievement

________ (b) had a generally beneficial effect

________ (c) had a generally negative effect

________ (d) allowed for better treatment of individual differences

________ (e) had a negative effect when used as the total method

________ (f) made possible dictation for homework

________ (g) had a beneficial effect when used along with other instructional methods

________ (h) other, please specify ________________

i. **Deal with only one topic in one question.** How can a teacher answer a question such as this: Do you favor teachers' active participation in state business education associations and AVA? Some might favor membership in one or the other of the organizations, but not both, so how can they answer the question? This matter should be handled in separate questions or in one question with a listing of the various organizations from which the respondent may choose.

j. **Remove emotional overtones from the questions.** Instead of asking "Do you favor abolishing the teacher-tenure system in public schools?" the question could be reframed in the form of questions about the advantages or disadvantages of the tenure system:

What do you feel are advantages of a nontenure system for teachers? Mark your choices with either 1, 2, or 3, with 1 representing the most important and 3 the third most important advantage:

________ (a) Keeps teachers alert to better teaching methods

________ (b) Keeps teachers from getting "rusty"

________ (c) Eliminates less competent teachers

________ (d) etc. etc.

________ (e) Other, please specify ________________________________

Such a question should be balanced with one asking about the disadvantages so that you don't give the impression that you are "pro" nontenure systems. Similar questions should also be included about the advantages and disadvantages of a tenure system. No questionnaire should give the impression that the researcher has "an axe to grind."

k. **Keep the response choices in mutually exclusive categories.** Be sure the respondent's reply fits into only one place or that a place is provided for any reply. For instance, a reader survey sent by a trade journal asks the reader to indicate the number of employees in the firm in which he is employed in an item like this:

Employees at this address: _____1-10_____11-19_____20-49
_____50-99_____100-249_____250-499
_____500-999_____over 1000

Where does someone who wanted to answer "1000" put his response?

6. **Prepare dummy tables of your anticipated responses.** Perhaps the best way to check whether your questions will provide usable data is to plan exactly how the data will be classified, tallied, and summarized into tables.

This should be an ongoing process as you develop your questions. Plan all tables completely before you even try out the questionnaire, even to the titles for the tables. This step serves as a double check on the adequacy of your questions. If the responses to a question cannot be made into a table, that is usually a sign that the question needs revision. Now is the time to overcome that problem rather than discovering it when the "real" responses come in.

7. **Design an attractive questionnaire.** As with any written communication, appearance of the instrument affects the reader's attitude toward it. A data-gathering instrument should reflect the work that has gone into it. It should be well arranged, and generally attractive. Several things may be done by the researcher to make an instrument more eye-appealing.

a. **Leave as much "white space" as possible.** A crowded questionnaire has a negative influence on a respondent—it just looks too complicated, as if too much work is required to read it and fill it out. This is especially true if the margins are narrow and very little space is left between questions. Two uncrowded pages may create more interest than one packed page that looks as if every effort was made to jam, cram, and stuff everything onto that one page.

If you have only two or three questions left for a second page, try folding the paper in half, making a four-page folder. Chances are you will be able to get all the questions on the one sheet of paper, with as much or more white space as you had with the one flat sheet of paper. An attractive folder is better than one flat, jammed, crammed, stuffed page. You are actually using both sides of the paper without its looking like a long questionnaire. If your questions are short, with short responses, the paper may even be divided into thirds for a very attractive questionnaire, with the questions printed across the narrowest dimension. Legal-size paper folded into fourths fits into a legal-size envelope and makes an attractive questionnaire. But use paper of a size that is easy to fit into envelopes and that is easy to handle and file.

b. **Make the questionnaire as short as possible.** Even though you try to keep the questionaire short, it must, of course, cover the topic and supply the information that you need. If the topic is of great interest and if the appeal to the respondent is well made, the length of the questionnaire may not be as important as its attractiveness. However, every effort must be made not to require too much of the respondent's time. The techniques discussed above for writing questions and for designing the format can all help to make a questionnaire seem shorter than it actually may be.

c. **Leave adequate space for the answers.** You are probably familiar with application blanks (which are really a form of questionnaire) that provide too little space for your previous work experience. Often the space is not adequate even for the firm name, let alone the address and the nature of your employment there. If you are asking for free response answers, do not expect the respondent to write between lines made by single-spaced underscores on a typewriter. In fact, use no lines at all—just leave enough blank space for even a moderately loquacious person to write responses, to encourage respondents to give more than a quick, superficial answer. In Question 17 below, the respondent is invited to make comments, but is allowed (on the original questionnaire) about an inch and a half of a single-spaced line for each one of them. This surely will discourage respondents from adding information that might be valuable in interpreting the responses.

> 17. Have you had to refuse students wishing to enroll in business education courses because you do not have enough funds?
>
> For grades 1-12: _____YES_____NO Comments:________________
>
> For technical schools: _____YES_____NO Comments:________________
>
> For four-year colleges: _____YES_____NO Comments:________________

d. **Line up responses under each other.** Having all the check marks or the written responses at the right margin facilitates data processing. Much time and effort are saved by not having to look all over the page for responses. If the researcher received two or three hundred responses to Question 16 below, the tallying or coding job would be a rather formidable task.

> 16. If you have follow-up information, approximately how many students are employed as business education teachers within a few months after graduation?
>
> _____0-9% _____10-39% _____40-59% _____60-79% _____80-100%

How much easier the task would be if the answers were aligned the way they are in this question:

What credits are required for the 6th year degree beyond the master's degree?

	Sem. Hrs.	Quarter Hrs.
a. Number of required credits in Business/Distributive Education		
b. Number of required credits in courses other than Business/Distributive Education		
c. Number of elective credits		
d. TOTAL number of post-master's credits required for degree		

If possible, the answers should be lined up at the right margin, since then, for most people, the question will not be covered with their right hand when they are marking their responses.

e. **Precode the responses if possible.** If you plan to use mechanical or electronic data processing, you can indicate the card columns beside each item number, and the choices can be numbered to indicate the code to be punched for the response. Careful planning of the coding while designing the questionnaire may save your having to transfer the data to code sheets for card punching. If you will refer to Illustration 3-3, p. 45, you will see that the question has been precoded for key-punching. The small number (46) beside Question 1 means that the data from that question will be punched in Column 46 on a punched card. A hole will be punched in the "1" position if the respondent indicates that the courses were more demanding in the summer than they were during the regular academic year, in the "2" position if the second response is selected, and the "3" position if the respondent chooses the third listed response. The responses for Question 2 will be punched in Column 47, and Column 48 will contain the responses for Question 3. With a format such as this, key-punching can be done directly from the questionnaire. The researcher may not have to prepare a code sheet for the key-punch operator to work from.

Sometimes more than one column must be allowed for a single item, as in this question:

3. Indicate the college in which you will earn your degree. (Col. 5-6)

Administrative Science	______01	Home Economics	______08
Agriculture	______02	Journalism	______09
Architecture	______03	Natural Resources	______10
Arts	______04	Nursing	______11
Arts and Science	______05	Social Work	______12
Education	______06	Other	______13
Engineering	______07		

Your major ______________________________ (Col 7)

Illustration 3-5. Precoding of Responses

For instance, since 13 choices are given for this question, each response requires two columns. A degree from the College of Natural Resources will be punched with a "1" in Column 5 and a "0" in Column 6. The response of someone in the College of Education will be punched with a "0" in Column 5 and a "6" in Column 6.

The free-response question asking the respondent's major will require hand coding as the questionnaires are returned, according to a classification scheme that will have been developed before the questionnaire is distributed. Perhaps the researcher had a list of possible majors from each of the colleges, which would be referred to in coding the responses. If, for instance, vocational education was coded 3 on the list of education majors, Columns 5, 6, and 7 would be punched 0, 6, and 3, respectively.

f. **Vary type styles and sizes.** With availability of typewriters that allow changing type styles and even size of type, variety can be introduced to add emphasis and to help organize the questionnaire. Headings may be done in a type style that stands out from the body of the questionnaire; important words may be italicized, typed in bold face, or, in the case of a typewritten questionnaire, underscored.

g. **Title the questionnaire.** Every data-gathering instrument should have a title that reflects the purpose of the study and should be more descriptive than "Questionnaire," or "Survey of Business Teachers." Below are some questionnaire titles from the writer's files of data-gathering instruments. Which are ones to which you would more likely respond? Which ones give you some idea of the contents of the questionnaire or of the purpose of the study?

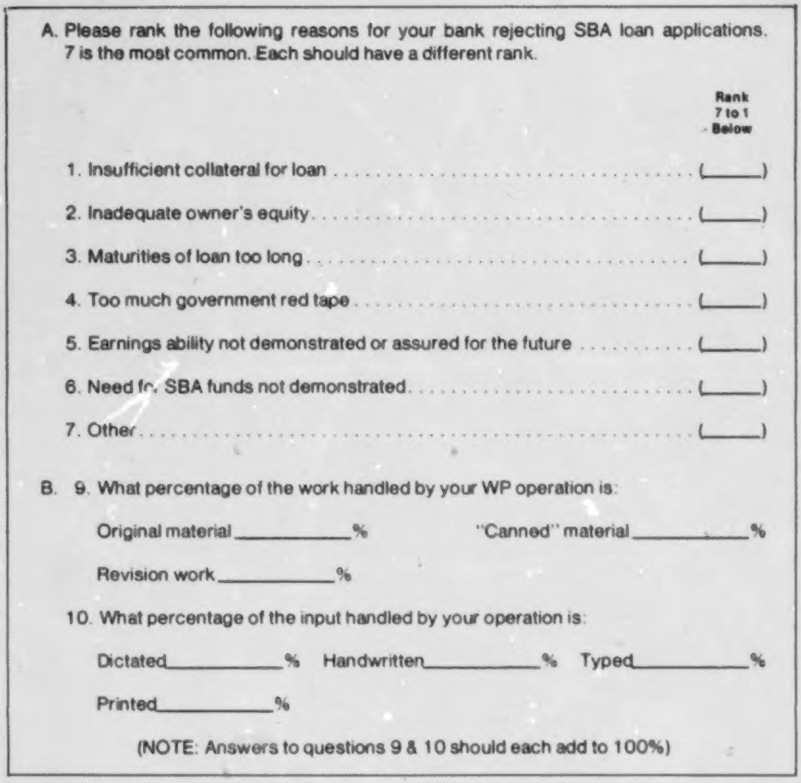
Survey
Survey of Instructional Programs in Data Processing
Questionnaire
Assumptions about People
Summer Session
Graduate Student Satisfaction Questionnaire
This We Believe about Business Education
Classroom Problems of First-Year Teachers

h. **Use professional quality reproduction.** Reproduce the questionnaire using a process that makes neat, legible copies. Using offset, Xerox, or a good stencil duplication job is far better than making the necessary copies of a questionnaire on a spirit duplicator. All copies, of course, must be clear, with dark, readable print, and free from smudges or other extraneous marks. The questionnaire must have the appearance of a high-quality, professional job.

8. **Supply clear, complete directions.** The respondent must be told exactly who should answer the questions and how the responses should be indicated. The respondents must know exactly what they are expected to do. Do they make a checkmark next to the choice that is their response? Do they mark only one choice or should they mark any that apply? Are they to rank from 1 to 5 their choices in order of importance? If so, they must know whether 1 or 5 will indicate the most preferred or highest ranking choice.

What form will the response take: a check mark? a letter placed in a blank? underscoring the response?

The directions are best placed wherever in the questionnaire they will be needed rather than all at the beginning. Also, sometimes an illustration of how to fill out a particular section is more helpful than a detailed statement of direction, as shown in the two samples in Illustration 3-6.



A. Please rank the following reasons for your bank rejecting SBA loan applications. 7 is the most common. Each should have a different rank.

Rank 7 to 1 Below

1. Insufficient collateral for loan . . . . (_____)
2. Inadequate owner's equity . . . . (_____)
3. Maturities of loan too long . . . . (_____)
4. Too much government red tape . . . . (_____)
5. Earnings ability not demonstrated or assured for the future . . . . (_____)
6. Need for SBA funds not demonstrated . . . . (_____)
7. Other . . . . (_____)

B. 9. What percentage of the work handled by your WP operation is:

Original material __________ %   "Canned" material __________ %

Revision work __________ %

10. What percentage of the input handled by your operation is:

Dictated __________ %   Handwritten __________ %   Typed __________ %

Printed __________ %

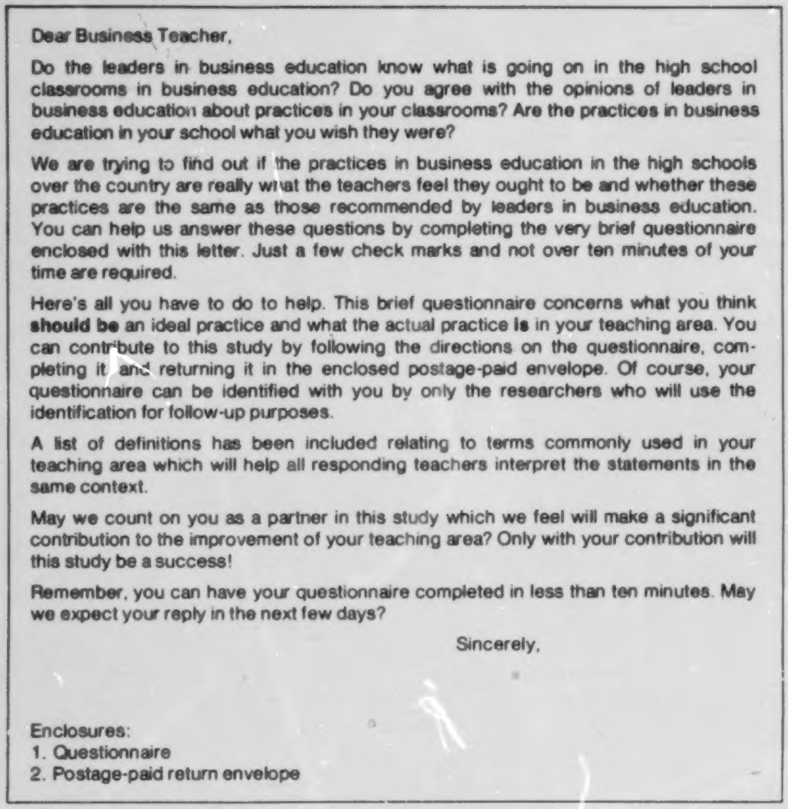
(NOTE: Answers to questions 9 & 10 should each add to 100%)

Illustration 3-6. Giving Explicit Directions for a Particular Question.

9. **Try out your questions.** No questionnaire should be mailed until it has been filled out by a small "try-out" sample of people. They should be asked to answer the questions in the most "far out" way they can, reading into your questions any misinterpretation they can. They should be encouraged to read the questions in every way they can be read, to see how clearly and unambiguously the questions have been constructed. These individuals should make notes and comments about anything they are not sure they have read correctly or are not sure they understand. Sometimes a graduate

class or a graduate seminar makes a good initial "proving ground" for the questionnaire. Then make any necessary revisions according to the suggestions that have been gathered before the questionnaire is sent to a small sample similar to the one from which data will be collected. Wording some of the questions two different ways may provide information about the ease with which they can be answered and the usefulness of the responses. This preliminary pilot sample should be asked to record the time required to answer the questions.

This try-out process should be repeated as many times as necessary to get the best possible questions. Only then is the instrument ready for use in the actual study.



Dear Business Teacher,

Do the leaders in business education know what is going on in the high school classrooms in business education? Do you agree with the opinions of leaders in business education about practices in your classrooms? Are the practices in business education in your school what you wish they were?

We are trying to find out if the practices in business education in the high schools over the country are really what the teachers feel they ought to be and whether these practices are the same as those recommended by leaders in business education. You can help us answer these questions by completing the very brief questionnaire enclosed with this letter. Just a few check marks and not over ten minutes of your time are required.

Here's all you have to do to help. This brief questionnaire concerns what you think **should be** an ideal practice and what the actual practice **is** in your teaching area. You can contribute to this study by following the directions on the questionnaire, completing it, and returning it in the enclosed postage-paid envelope. Of course, your questionnaire can be identified with you by only the researchers who will use the identification for follow-up purposes.

A list of definitions has been included relating to terms commonly used in your teaching area which will help all responding teachers interpret the statements in the same context.

May we count on you as a partner in this study which we feel will make a significant contribution to the improvement of your teaching area? Only with your contribution will this study be a success!

Remember, you can have your questionnaire completed in less than ten minutes. May we expect your reply in the next few days?

Sincerely,

Enclosures:
1. Questionnaire
2. Postage-paid return envelope

Illustration 3-7. A Cover Letter for a Questionnaire

EACH QUESTIONNAIRE COUNTS . . .

Yes, even though the response to the questionnaire has been most gratifying, we are still anxious to receive your completed form. You may be interested to know that as of November 16, 60 percent of your fellow business teachers throughout the nation have returned their completed questionnaires.

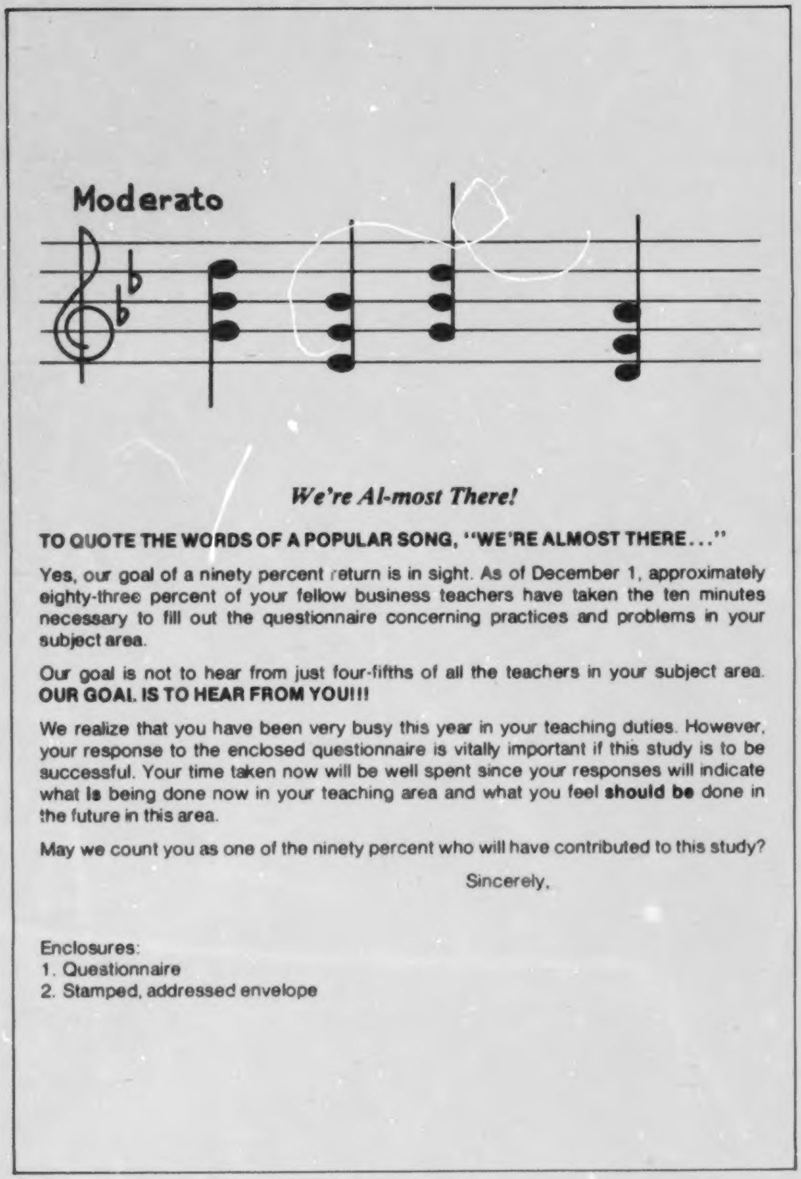
In order to have a valid representation of secondary school business teachers in this nation-wide study, a return of approximately 90 percent is needed. Won't you help us reach this goal? You can do so by completing the enclosed questionnaire and returning it in the postage-paid envelope.

You can complete your questionnaire in such a very short time. May we hear from you immediately?

Sincerely,

Enclosures:
1. Questionnaire
2. Postage-paid envelope

Illustration 3-8. A First Follow-up Letter



Moderato

We're Al-most There!

TO QUOTE THE WORDS OF A POPULAR SONG, "WE'RE ALMOST THERE..."

Yes, our goal of a ninety percent return is in sight. As of December 1, approximately eighty-three percent of your fellow business teachers have taken the ten minutes necessary to fill out the questionnaire concerning practices and problems in your subject area.

Our goal is not to hear from just four-fifths of all the teachers in your subject area. **OUR GOAL IS TO HEAR FROM YOU!!!**

We realize that you have been very busy this year in your teaching duties. However, your response to the enclosed questionnaire is vitally important if this study is to be successful. Your time taken now will be well spent since your responses will indicate what **is** being done now in your teaching area and what you feel **should be** done in the future in this area.

May we count you as one of the ninety percent who will have contributed to this study?

Sincerely,

Enclosures:
1. Questionnaire
2. Stamped, addressed envelope

Illustration 3-9. A Second Follow-up Letter

10. Writing a good cover letter and follow-up reminders. An appropriate appeal in the letter that accompanies the questionnaire will help greatly in increasing the response rate. The cover letter is a sales letter. It should sell the importance of the study, ask the cooperation of the individuals, and make them feel their contribution is of real value to the study, which in turn is a contribution to business education. Visualize a typical recipient and then sell the study on the basis of what you feel will be most appealing to such an individual. Also include an estimate of the time required to fill in the questionnaire and give assurance of the confidentiality of the responses.

Usually an altruistic appeal, what the respondents will be doing for the field of business education, will elicit as much or more cooperation than material incentives. One appeal that does not motivate most respondents is that you are doing your study as part of your advanced degree. Your study should have enough value to stand on its own merits!

The letter in Illustration 3-7 is from a study in which a response rate approaching 90 percent was achieved in a study involving a sample of 1,500 high school business teachers. Notice that the letter begins with an attention-getting paragraph. The questions raised in it really summarize the study. The respondent is made to feel the questionnaire is easy to fill out, and the whole letter sounds as if it is written to one individual.

Illustrations 3-8 and 3-9 are two of the three follow-up letters sent two and four weeks after the original mailing of the questionnaire. Even though these letters have a rather light approach to the follow-up, they each brought in about 25 percent more questionnaires. One more follow-up was sent, resulting in an 88 percent return in the study. Other researchers may prefer a more serious approach in their follow-up reminders, but in either case the follow-up letter as well as the cover letter must sell the respondent on the value of the study and of their contribution to it.

Interviews

Some of the advantages and disadvantages of interviewing as a means of gathering data were implied in the section on the use of questionnaires: Interviews allow more in-depth probing into the answers given by the respondent than does a mailed questionnaire; questions may be clarified; observations may be made of the respondents' attitudes toward the study that may be influencing their responses; and in certain sensitive areas of questioning, a good interviewer may get a higher rate of response than one can with questionnaires.

Drawbacks of interviews include the greater cost per respondent; limitations on the number of persons that can be reached; "error" due to interviewer variability—for instance, the attitudes they may exhibit (perhaps unconsciously), personal beliefs that may subconsciously enter their "probing," or the rapport they are able to establish with the respondents.

The same careful planning is necessary for gathering data by interviews as by questionnaires. An additional decision, however, is how structured or open the interview will be. Unstructured interviews in the hands of an unskilled interviewer can elicit a great deal of "chaff" and may miss some of the points that should be covered. Unstructured interviews require thoroughly trained, skilled interviewers. On the other hand, unstructured interviews may allow for greater freedom for probing, for getting to the "real" responses.

Even if most of the questions are structured and are to be asked the same way by the interviewers of all the respondents, the individuals gathering the data must learn to recognize when interviewees seem to misinterpret a question and when they may be giving incorrect data for some reason. The attitude of the interviewer and the rapport established with the interviewee can influence responses in both structured and unstructured situations.

Unless you have extensive funding for the project and are knowledgeable about interviewer training, you should probably avoid interviews unless you can do your own interviewing—and then proceed only after very careful and extensive practice interviewing.

Observations

Some types of information cannot be obtained by questionnaire, by testing, or by interview; namely, when the investigation is concerned with actual behavior of subjects that they cannot themselves report accurately. Let's say that of primary interest are differences in achievement and attitudes toward business law in classes in which students are encouraged to express themselves frequently and freely and in classes in which the teacher does most of the talking. Obviously the teachers cannot be asked to estimate how much of the time they talk and how much the students talk. In order to get reliable information about the ratio of student/teacher talk, observation of what actually goes on in the classrooms is called for.

Problems in Using Observations

Observation as a data-gathering technique has some of the same advantages and limitations that interviews have in that the observers need to be trained. They must be trained in exactly what to observe and how to record their observations. Considerable practice is necessary for individuals to learn these techniques. Among other problems in using observation as a means of getting data are these:

1. **Personal biases of the observer.** Subconscious though they may be, biases of the observer can influence the observation. A common bias in such situations is the halo effect; that is, the general impression gained from the first observation or two may influence how the observer feels in general about that particular classroom, and those feelings may carry over to future observations.

2. **The background the observer brings to the situation** may affect what is seen. For instance, a relatively authoritarian teacher sent to observe in an open classroom is likely to see and interpret events in the classroom very differently from another teacher who has had experience with an open classroom situation.

3. **The degree of inference required.** The greater the degree of inference between the behavior observed and the characteristic being studied, the greater the observer error is likely to be. Let us assume a situation in which data are being gathered for the purpose of inferring a characteristic of teacher confidence. Various actions forming the behavior of the teacher, things said, nonverbal behaviors, physical gestures, all may be the events recorded by the observer as evidence that the teacher either has considerable confidence in his teaching or lacks confidence. The greater the degree of inference required, the more room for observer error in selecting items to record or in interpreting events that take place.

Often one must infer a condition from a behavior; for instance, attendance, attention in class, and willingness to do necessary practice may be evidences of motivation of the students. The inference is made that if students display such behaviors, they are motivated. Looking attentive in class may only be "play acting" on the part of the student; thus, if the observer records it as evidence of motivation, we have observer error.

4. **Personal experiences of the observer.** Much subjectivity enters into observing, especially in observing complex behavior. Again, experiential background of the observer may influence how much and what should be recorded and what the observer considers to be significant behavior. Witnesses to an accident, for instance, "see" very different things and almost as many versions of the incident may be reported as the number of witnesses.

5. **The presence of the observer.** Some persons feel that the fact that the observer is in the room may influence the behavior of the individuals being observed. However, as Kerlinger (1973) points out, this may not be as important as some seem to think. After the person who is observing has been in the room a few days, the people involved seem to disregard the observer. Also, even though everyone may "put his best foot forward," Kerlinger points out that "one cannot do what he cannot do" (p. 539).

6. **The observer's prior knowledge.** If the observer knows what behavior patterns are likely to occur in the particular situation, this knowledge may influence the observation. Contamination may occur if the observer knows how the subjects may behave because of something he knows about the subjects. This was a possibility in Hansen's (1967) study of behaviors of high-creative and low-creative teachers in basic business classes had she not guarded against it. Because high-creative teachers may have had more novel and a greater number of approaches to the subject matter than the low-creative teachers, Hansen had a university research bureau supply her with the names of the six basic business teachers who had the highest and the six

who had the lowest scores on a creativity test that the bureau had scored, with no designation of any teacher's score. If the observer knows something about the subject being observed, the observer may "predict" the subject's behavior.

Precautions in Making Observations

The shortcomings of gathering data by observation may be partially overcome by taking certain precautions. The observers may be supplied with observation schedules that guide them and restrict their observations to specified behaviors. Several well-known classroom observation scales are available, but their use requires training and practice (see Kerlinger (1973) for names of these instruments). Limiting the observations to one or two specific kinds of behaviors or events greatly increases the reliability and validity of the observation data.

Accuracy of observations may be increased by use of recording devices such as audio tape recorders or videotape recorders. Observers can then replay the tapes as many times as necessary in order to record everything they may wish to record in the situation being observed.

Then, of course, the degree of agreement among observers must be determined and reported. Observers should be trained, and they should practice making observations until they show evidence that they are "seeing" the same things in the same way. Hansen and one other observer used a modification of the Flanders Interaction Analysis Scale in the study of high- and low-creative teachers. They practiced recording every 30 seconds what was being said by either a student or the teacher or any silence that occurred at the time. They had over a 90 percent agreement on the classifications of the talk-behavior they observed.

Task Analysis

A common form of observation in vocational education is task analysis. The data may be recorded by the worker, or someone may observe employees at work and record their activities either in narrative form or on a recording sheet. Task analysis as a basis for curriculum decision-making has some shortcomings, however. These are mainly the difficulty in evaluating the relative importance of the task and the place it should hold in the curriculum. One problem is that of selecting from among the tasks those for which the students should be provided instruction. For instance, some tasks performed by typists (addressing envelopes, as an example) occur quite frequently, and so occupy quite a high position in the list of typists' tasks arranged according to frequency. On the other hand, a market analysis report may be typed only four times a year. However, its preparation is a great deal more complex than addressing envelopes—it may be a more "important" task than envelope-addressing. Thus, the frequency with which a task is performed is not an adequate basis for deciding the emphasis to be given in the office education curriculum to preparing for certain tasks.

Another aspect of preparing students for performing on the job usually not considered in task analysis is the concommitant knowledges and reading levels required for performing satisfactorily on various jobs. Task analysis may pay some attention to knowledges, but very few if any studies have considered the level of reading skill required an element of the job.

The matter of developing a means of weighting the relative contribution of frequency and importance is under study by several groups. One promising task analysis technique is that developed by the Department of Labor. It is explained and instructions given for performing such task analyses in the *Handbook for Analyzing Jobs* (United States Employment Service, 1972).

Observation is a valuable, promising technique, if one makes every effort to overcome the problems of observer reliability with some of the suggestions offered here. It, like all other data-gathering techniques, is only as good as the care taken in planning and carrying it out.

Attitude Measurement

Many studies are concerned with the attitudes of individuals toward certain practices, events, or ideas pertaining to business education. Usually one is interested not only in whether the subjects are favorably or unfavorably disposed toward the concepts under study but also in the strength of their feelings, as well as the particular aspects of a concept they may or may not favor.

Likert Scales

Perhaps the most widely used means of determining attitudes is a Likert scale. A Likert scale is a series of statements with which the subject is given the opportunity to agree or disagree and to indicate the degree to which he does so, usually on a five-point scale. These scales are very widely used because they allow one to quantify attitudinal data by assigning numerical values to the responses made by the subjects. An example of the type of item appearing on such a scale is this:

1. Basic business classes should be taught in the twelfth rather than the tenth grade

SA A ? D SD

Or it may take the form

2. Basic business classes should be taught in the twelfth rather than the tenth grade

|_____|_____|_____|_____|

Strongly Agree — Strongly Disagree

Values from 1 to 5 are assigned to the 5 positions, with Strongly Agree usually receiving a value of 5. Some of the items should also be stated negatively, in which case the 5 would be assigned to SD, since disagreement with the statement indicates a positive attitude. For instance, one may infer a positive attitude toward vocational education from a negative reaction to this statement:

> Vocational education occupies too prominent a place in the high school curriculum.

The sum of the values assigned to each of the statements comprise the subject's score and is indicative of his attitude toward the concept being studied. The means of the attitude scores of various groups of people can be compared, or the mean scores of a single group before and after studying a particular unit or course may be compared. The lowest score possible on a 25 item instrument, if a subject Strongly Disagreed with every statement, is 25, while the highest score possible if someone marked SA for each of the 25 items, would be 125; a completely neutral person (if such a person exists) would have a score of 75.

One must keep in mind, however, two important considerations in the use of scores derived from Likert scales; namely (1) the importance of the subject's familiarity with the concept, and (2) that identical scores may arise from entirely different sets of attitudes. One individual may strongly agree and agree with one set of statements and another person may strongly agree and agree with the same number of statements, but an entirely different set, yet both respondents would have the same total score. However, if one is comparing groups, this is not too serious a matter, because one is looking mainly at the overall picture of attitude of the group toward the topic rather than individual attitudes.

Another limitation to consider is that individuals tend to respond to all items in a certain way; for instance, they may never use the SA or SD choices, or they may tend to mark all statements "agree" whether or not these are their real feelings. To overcome such occurrences, one must be sure that the sample of persons asked to fill out the scale is from an appropriate population—persons who are interested in and knowledgeable about the topic so they have an adequate frame of reference for making the judgments they are asked to make. Another procedure to eliminate choice preferences is to use both negative and positive statements in random order, so no pattern is apparent in the items.

Attitude statements should be arranged from general to specific and should be free from suggestion of what the response should be. The number of scale divisions is usually 5, but could be as few as 3, and usually should not be more than 7, since most individuals cannot distinguish clearly among the middle range of choices if given more than 7 categories.

A Likert scale is a useful technique if the instrument is carefully constructed, with pretests made of it to establish validity and reliability. The technique most commonly used to measure reliability is that of determining the internal consistency of the items. This is determined by correlating the score assigned to each statement by each individual and that individual's total score.

The following example of a tally may help clarify this procedure:

Statement #1

Subject	Statement Score	Total Score
1	3	65
2	5	100
3	2	90
4	2	65
.	.	.
.	.	.
n	5	105

A coefficient of correlation (r) is calculated between the statement scores and the total scores. A similar set of two distributions is made for each item on the instrument. Items with the lowest correlation coefficients (lowest reliabilities) can be eliminated from the final form of the instrument or they can be revised and the instrument can be retested with another sample of subjects. Validity may be established by administering the instrument to groups whom the researcher has good reason to believe hold a certain view of the concept about which an attitude is to be measured. Devising one's own Likert scale can be a time-consuming, relatively complex procedure, requiring good knowledge of measurement theory. However, one who plans to use one of the available attitude scales should have some understanding of how they were developed and be sure that they actually do measure the attitudes one wishes to study.

Semantic Differential

One of the main limitations of gathering data about attitudes by questionnaires or attitude scales is that respondents can often "fake" the responses—they can answer the way they think the researcher wants the items answered or in a way that puts them in what they think may be a more socially acceptable light than their true feelings might. If the questions are of a personally sensitive or threatening nature, the respondents may be reluctant to answer at all. A promising, more subtle approach to measuring attitudes, the Semantic Differential, has been developed by Osgood (1957). The Semantic Differential measures the connotative meanings of concepts in what Osgood calls semantic space, which has dimensions just as a box or a room has dimensions. We can describe a box as having length, width, and height, 6 units long by 2 wide by 4 units high and another box with dimensions of 4 by 1 by 6; we know that the second box is shorter, narrower, but taller than the first box. The second box has more of that characteristic (dimension) called height and less of width and length. So it is in measuring the perceptions of a particular concept held by two groups of people. The two socially different groups (for instance, union members and the management team of a corporation) are likely to have different feelings toward certain issues or concepts related to management-labor relations. Management may feel strongly that right-to-work legislation is desirable; on the other hand, union members may tend to feel strongly that such legislation is not good.

These particular feelings illustrate two of three dimensions of semantic space identified by Osgood; namely, potency ("toughness" of feelings) and evaluation (a quality such as goodness or badness). The third dimension is labeled activity (expressing motion or action). Other researchers have developed other dimensions, but the original three seem to be the ones most widely used in educational research.

Basically, a semantic differential consists of a series of concepts related to a topic or issue (objects, practices, traits, etc.), each of which is placed at the top of a separate page or card. Listed below each concept is a set of adjective pairs, usually seven to eleven pairs, one adjective of the pair having a positive and the other a negative connotation. The two adjectives are placed at either end of a line which is divided into seven equally spaced segments, as in Illustration 3-10. Osgood has validated a set of 53 adjective pairs, each of which measures one of the three dimensions more strongly than the others. One then can select adjective pairs from the list that may seem appropriate to measure the particular dimension(s) of the concept about which one is concerned.

Historical Development of Vocational Education

Pleasant :____:____:____:____:____:____:____: Unpleasant

Awful :____:____:____:____:____:____:____: Nice

Cruel :____:____:____:____:____:____:____: Kind

Rugged :____:____:____:____:____:____:____: Delicate

Good :____:____:____:____:____:____:____: Bad

Slow :____:____:____:____:____:____:____: Fast

Honest :____:____:____:____:____:____:____: Dishonest

Unfair :____:____:____:____:____:____:____: Fair

Illustration 3-10. A Semantic Differential Item

Each respondent is asked to indicate on the seven-point scale his reaction to the concept in terms of each of the pairs of adjectives on the card. Participants in a study are likely to have different feelings to a greater or lesser degree for each adjective pair for each concept, and they mark different places along the scale to indicate their feelings and the strength of their feelings regarding each adjective pair. Numerical values are assigned to the positions on the scale, and thus for each concept a value for each dimension of the concept for that respondent may be calculated as in a Likert scale. Means of a person's responses to each concept then can be calculated to measure his feeling toward the particular concept. Adjectives may be selected for only one dimension if that seems more appropriate; most commonly the evaluative dimension is the single aspect of a concept that is studied.

The values for each concept in the scale can then be added to get a measure of a person's attitude toward the particular issue. One can also calculate the mean attitude score for one group to compare with the mean of another group.

Critics of the technique point out that the adjective pairs may not seem to be at all related to the concept. However, Osgood and others (Kerlinger, 1973; Tuckman, 1972) point out that the irrelevance really seems to be the strength of this approach since according to Tuckman "it often limits the tendency to produce socially accepted responses" (p. 161).

The use of this technique is perhaps most clearly illustrated by describing a study investigating the attitudes of basic business teachers toward the use of selected teaching activities and media. Since attitudes may influence actions, methods teachers should know whether or not certain psychologically sound techniques are accepted by teachers. If not, perhaps more emphasis on and practice in their use in methods classes may be appropriate.

A pilot study was done first in which the researcher selected the evaluative adjective pairs with the highest correlations between them and a criterion score reported by Osgood by which a group of teachers could express their feelings on 12 concepts. All adjective pairs were eliminated for which the mean value of the group was near the neutral position, and the final instrument contained only those adjective pairs to which strong responses were made.

Data were analyzed according to the amount of experience the teachers have had, their preference of teaching field within business education, the size of school in which they teach, etc. Asking teachers whether or not they favor games or simulations or lecturing, discussions, etc., in basic business classes in straightforward questions about these techniques may have elicited responses likely to be what respondents considered to be "right" answers rather than an expression of their true feelings. In this case the more subtle approach of a semantic differential was deemed necessary to increase validity of responses.

Q-Sorts

Another interesting technique may be used to explore changes or differences in opinions or self-concept; namely, Q-Sorts. This technique also has been used to study differences in esthetic judgments and preferences. Even though it is generally more time consuming than answering a questionnaire or marking Likert scales, respondents are likely to cooperate because it is an interesting technique.

Respondents are asked to sort a series of statements, each one written on a small card, into a designated number of piles, each pile to contain only a specified number of cards. The statements may be statements of characteristics which the subjects sort into piles ranging from "most like me" to "least like me;" or they may be statements of behaviors that range from "most effective" to "least effective" or "most favorable" to "least favorable."

To illustrate one possible use of this technique, consider the Ford and Patterson (1967) study of the problems of student teachers as viewed by student teachers, college supervisors, and the classroom cooperating teachers with whom the students worked.

Teachers and business education graduate students were asked first to list all the problems they could that they had observed student teachers having. This list was reduced (combining similar problems, eliminating duplications) to a set of 54 problems, each of which was written on a small card. Packets of these 54 cards were distributed to each student teacher, to the student's college supervisor, and to each of the classroom teachers with whom each student worked.

Each individual was asked to sort the 54 cards containing student teaching problems into five piles, representing degrees of seriousness (or typicalness) of the problem for the designated student teacher, with Pile #1 being the least typical or least serious problem. The sorters were instructed to sort and re-sort until they had a prescribed number of cards in each of the piles, with the distribution of cards resembling a modified normal curve.

Numeric values are assigned to the piles, ranging from 1 to *n*, the number of piles used in the study. Usually the number of piles will be anywhere from 5 to 11. The greater the number of statements (recommended number is between 60 and 120), the greater the number of piles into which they may be sorted. Any statement placed in the "least typical of me" pile would receive the lowest value and the other extreme would be assigned the highest score.

In the Ford and Patterson study, the mean score was calculated for each of the three groups of individuals for each of the 54 statements, and the statements were placed in rank order according to the mean value for each group. Thus the rankings of the three sets of 54 rank-ordered statements were compared.

This technique could be used in different ways: One could measure attitudes toward office work, for instance, held by students in different kinds of office practice programs or measure differences in attitude at the beginning and end of an office procedures course. Or one could have groups of college students with different majors sort statements about the American economic system.

74

In some studies the cards are sorted with the same number of cards in each pile, and in others no specification is made about the number of cards to include in each pile. This last procedure is utilized in establishing items for the Delphi technique for arriving at a consensus.

The Delphi Technique

In certain types of research, consensus of experts is required, perhaps as a basis of developing criteria of some sort, for instance; but getting the experts together to arrive at a consensus is often impracticable, in terms of both time and money. Conferences for the purpose of arriving at decisions about procedures, recommending courses of action or determining characteristics of good teachers, for instance, often are inefficient and ineffective ways of reaching agreement on these matters. The Delphi technique is a means of arriving at a consensus without bringing the panel together in a face-to-face situation. It was developed at the Rand Corporation (Helmer, 1967) about 20 years ago for the purpose of making forecasts.

In group situations many factors influence each individual's decision: his own status in the group; the influence of others, perhaps even the loudness and persistence of one or a minority of the group; or even an unwillingness to change one's mind from publicly expressed opinion. Proponents of the Delphi technique claim it eliminates these problems of group activity, employing instead a series (usually three rounds) of questionnaires or opinion scales sent to individuals, interspersed with feedback regarding opinions and information provided by the other participants. The panel members, however, are never told who the other participants are, so they will not be influenced by such knowledge. The technique is perhaps best explained by briefly describing two studies in which it was used in two different ways for different purposes.

In one study the researcher (Chaney, 1972), who wished to develop a rating scale to evaluate student teachers in skills subjects in business, used the Delphi technique to establish the competencies that should form the basis of such a rating scale. The researchers in another study were interested in determining the opinions held by various groups for assessing the needs, desires, and opinions of the clientele of a school of education (Cyphert and Gant, 1970).

Chaney solicited the cooperation of seven nationally prominent business teacher educators to whom a series of three questionnaires was sent regarding competencies needed by skills teachers. The first-round questionnaire consisted of 47 items derived from competencies delineated in the literature as being desirable ones for skills teachers. The panel members were asked to indicate whether they thought each of the competencies was essential, important, or unimportant. They were also asked to add any other items they felt should have been included, as well as to make revisions or additions to the items on the instrument.

The second-round questionnaire gave the results of the first questionnaire, with the particular respondent's response indicated by a star, as in Illustration 3-8. If the respondents were in the minority, they were to reconsider the response and either change it, or, if they wished to remain in the minority, to indicate the reason for doing so. Five new items had been added to the instrument by the panel members, making a total of 52 characteristics to be examined on the second-round questionnaire.

	Essential	Important	Unimportant
Makes provision for individual differences	6*	1	
Gives "slow-motion" demonstrations	1	3	3*
Gives expert level demonstrations	4*	3	
Prepares bulletin boards and teaching aids		6	1*

Illustration 3-8. Panelists' Response Pattern on Questionnaire I of a Delphi Procedure, Indicating Respondent's Position

A third questionnaire was then sent, with the responses from Questionnaire 2 indicated, and the procedure was repeated, except no additions were to be made. Following the series of questionnaires, the characteristics rated "essential" by a majority of the panel members were incorporated by the researcher into a rating scale that contained 37 items. Some of the items were combined and/or slightly revised for the rating scale; for instance, the scale contains the item "Demonstration of the technique being taught" instead of the two items shown in Illustration 3-8 regarding demonstration in "slow motion" or at the "expert level." Incidentally, all five items suggested by the panel appeared on the rating scale. Chaney then further validated the student-teaching rating scale by correlating the responses of the college supervising teachers with the responses of the cooperating teacher in evaluating student teachers in classrooms.

While Chaney used the technique to determine what should be included on a student-teacher rating scale, Cyphert and Gant (1970) used the Delphi technique to determine opinions held by various groups toward a college of education. In this case opinions were solicited from seven groups of persons, including some professional educators (teacher-educators and administrators), some lay groups (involved with PTA, school boards, etc.), public officials, civic groups (labor, farm, newspaper editors, etc.) as well as some classroom teachers. The first questionnaire in the study was to determine those "prime targets on which the School of Education should concentrate its energies and resources in the next decade" (p. 418). Instead of using literature or other sources, as Chaney did, to develop items on Questionnaire I, the first round was used to solicit the items for subsequent questionnaires.

In contrast to Chaney's study, the respondents in the Cyphert and Gant study were not all "experts," and some refused to participate on the grounds that they lacked expertise in the field. Cyphert and Gant pointed out that "in order for prospective participants to take part in a study, they must be made to feel that their response is valid" (p. 421). The original sample consisted of 42[illegible] individuals, of whom 298 responded, and of whom, in turn, 262 responded on the third round. Lay persons as well as educators were reported to have considered the Delphi technique interesting and to have requested more information about its use. They indicated that, although considerable time was required to complete each of the questionnaires, they did so because of the importance of the study and their interest in the technique.

Sackman (1975), however, has raised some important questions about the Delphi technique of which a prospective user should be aware. Among the first criticisms is one in which he is concerned that the panel of experts is usually neither of adequate size nor does it fulfill the assumptions of randomization or of independent measures required for statistical analysis. The panel of experts tends to consist of persons whom the investigator knows or who have been recommended as participants because such persons are likely to cooperate and not drop out of the study. Thus, the sample of experts is likely to include only those persons with similar backgrounds and interests, who think along the same lines, and/or who have a vested interest in the area under investigation.

Another criticism leveled at the Delphi procedure by Sackman (1975, Chap. 5) is that in such a study, no one is accountable for the results and their use as a basis for decisions: the investigator objectively reports expert opinion; the panelists are protected by anonymity; and the user of the findings says that he is following the best advice available from experts and he is not responsible for what they say. However, the same may be said of any survey of "expert" opinion, whether by questionnaire, Q-sort, or any other means of soliciting their opinions. The soundness of the responses depends strictly upon the veracity and knowledgeability of the respondents.

Of more serious consideration, however, are Sackman's concerns about the influence of the feedback received by the participants after each round. After the first round the judgments are no longer independent; the second and third rounds "produce strict correlated or biased judgments" (p. 15). Since on each round each panelist is usually given information regarding his position in relation to the others, this lack of independence elicits concern by Sackman that all the reasoning and logic concerning the panelists' reconsidering their selections, their incorporation of new information, and their converging toward consensus "cannot hide the fact that independent judgment is destroyed once the participant knows how others have responded to each item" (p. 16).

Thus, since the assumption of independent judgment no longer holds, the standard statistical methods for testing hypotheses are no longer suitable for drawing inferences.

Neither does the Delphi method eliminate the influence of group pressure on the respondents' opinions. Sackman points out that real consensus is group agreement reached by the mutual education of the participants through information they gain through the adversary process. This, in turn, leads to improved understanding and insight into the issues, but "it does not refer to changes of opinion associated primarily or exclusively with bandwagon statistical feedback" (p. 48).

Sackman further says that the Delphi technique "deliberately manipulates responses" (p. 51) to reduce differences among the panel members, claiming this to be consensus. He says further that after the third or fourth round the individualist, the one who maintains his position, is a threat in that lack of agreement demands yet another round of the questionnaire.

These are serious criticisms and anyone contemplating using the Delphi method for getting consensus will do well to consult Sackman's book.

In the Delphi technique, as in any of the other data-gathering techniques described in this chapter, one must keep in mind that surveys of **opinion** do not form an adequate basis for policy decisions. That is not to say that surveys of opinion are of no value; they are for certain other problems and when they are supported by facts. But surveys and all other data gathering must result in valid and reliable findings, whether they serve as bases for further hypotheses to be tested in an experimental setting or for exploring the state of the art.

Chapter 4

Experimental Research

As you will recall from the first chapter, the element of a research design that makes one study an experiment and another an associational study (even though the same statistical analysis may be made of the data) is that the researcher actually manipulates the independent variable(s) in an experiment. Also, you recall that manipulate means either (1) actually to change circumstances under which the subjects operate (e.g., two groups of students are taught by two different methods, or groups of students are given different types of reading assignments), or (2) to assign students randomly to particular treatments (e.g., to measure the effect on learning of the presence or absence of windows in the classroom, students are assigned randomly either to a room with windows or to one without).

In associational studies, on the other hand, the students are grouped according to some pre-existing condition—they may be children of parents of given ethnic groups, or children of specified chronological age to be given certain physical tests. The students cannot be assigned randomly to parents of a particular ethnic group—the subjects' ethnicity is a pre-existing condition, as is his age; and subjects are selected for study because of some pre-existing characteristic. Children cannot be assigned to an age group; they are that age, and that is why they were selected for study.

The main objective of experimental research is to be able to predict, to attribute cause to effect. "What will happen if..." or "Under this condition of learning, we can be sure (with a specified degree of probability) that..." That is, we can be sure with the specified degree of confidence (arrived at by statistical analysis) **only if the study is designed so that anything that we have not accounted for is likely to have influenced the results of the study.** For instance, one reads about a teacher who experiments with a new method reporting increased skill in doing addition problems in this year's business mathematics class. The teacher finds that, after the introduction of this new approach, the students can do in one minute five addition problems each consisting of five numbers of three digits each. This may be a pretty good standard for this teacher's class, but is it better than what classes taught some other way may accomplish? How much of the class's achievement came from the method used and how much because the students were brighter than the teacher's previous students? Properly designed experiments help to provide an answer.

Designing Experiments

Experiments in business education take many forms. Some may require only a simple design, while others call for a complex design where many things need to be measured and controlled. Of primary importance in experimentation, though, is use of a design that fits the particular research situation, based on the particular hypotheses to be tested and the necessity for controlling variables. The planning and design must be decided on BEFORE collecting data, however. In the following pages, the interdependence of data collection, hypotheses, and statistical analysis will again be demonstrated.

The most typical of the experimental designs used in business education studies are presented in the following pages. However, the researcher should not attempt to force the data or experiment to fit into one of these designs if it does not result in the kind of information being sought. Many variations of these basic designs are possible. A statistician can provide invaluable help in explaining and planning the more elaborate designs and their analysis.

One-Shot Case Study

Teachers in a department may say that one of their colleagues is experimenting because that teacher is using a new approach to teaching addition in that one section of business mathematics. But is what the teacher is doing an experiment? In that the teacher used a new procedure (manipulated the independent variable) some might consider this an experiment. Campbell and Stanley (1966) label such a procedure the "**One-Shot Case Study,**" which they say is not a true experiment. Let us see why. Consider some of the things that could have accounted for the seemingly good performance of the students:

1. The teacher's enthusiasm for the new procedure could have motivated the students to do a better job.
2. This year's students may have been a more select, brighter group than previous groups.
3. The students may have been more motivated for some reason—a group who just liked business math.
4. The teacher's memory about the accomplishments of the previous year's class may have been faulty.

Furthermore, we do not know whether the class is really any better at adding at the end of the time than they were before the new procedure was introduced. Thus, introducing an experimental variable without a comparison group or without any specific prior knowledge about the group does not allow us to evaluate the effectiveness of a procedure.

Several other things could have been listed that might have accounted for the results, but these are enough to illustrate the lack of certainty we are likely to have in such a "study." A study using this design cannot attribute the result to the experimental procedure, nor does it give us any indication that it is likely to be effective with another group of students. This design may be depicted symbolically this way: $T_a O_{1a}$, which says that a treatment

(T_a) was introduced into a group, after which an observation (O) was made of the result.

One-Group Pretest-Posttest Design

Another teacher is interested in the effect of certain types of drills on straight-copy speed in typing, so the teacher gives the students a series of three-minute timings and records the means of the students' gross rates on these timings. The special drills are then introduced for a period of two weeks, after which the students type the three-minute timings again.

This design, known as the **One-Group Pretest-Posttest Design,** (Campbell and Stanley, 1966), may be depicted this way: O_{1a} T_a O_{2a}, which says that an observation (O_{1a}) was made of one group (a), which was then given a treatment (T_a), after which a second observation (O_{2a}) was made of the group's performance. Such a procedure does tell us whether or not the students were more skilled after having been exposed to the new drills than they were before, but it still doesn't tell us much. We do not know if the students' skill might have increased that much by using the drills they had been using prior to the special drills or if some other drills might have been as effective. Again, the teacher's enthusiasm, the students' prior abilities and/or motivation, the type of material they typed, the time of year, or many other things could have accounted for the gain. The One-Group Pretest-Posttest Design, too, is not a true experiment.

Such designs are commonly referred to as action research, and teachers should be encouraged to use them **provided they do not draw conclusions about the value of the procedure if it is applied to another group.** Action research is helpful for finding out whether a new procedure is feasible or whether the students do or do not seem to like it, etc. But no recommendations should be made about the general effectiveness of the "new" procedure on the basis of these designs.

In some pretest situations, whatever change resulted from a procedure may actually have come from effects of the pretest itself, as, for instance, in a study of the effect of certain instructional materials on attitudes toward certain topics in consumer economics. Some of the items on an attitude scale given as a pretest may alert students to something that may particularly interest them and thus create an influence on their study of the particular unit in a way that might not have occurred without the pretest.

Time-Series Design

Since the design with a pretest does not really give us information about the effectiveness of the treatment, some investigators will make a **series** of measurements before introducing the new procedure and then take a **series** of measurements following the new procedure. For instance, a shorthand teacher wishes to try a new incentive plan for encouraging the students to write more homework practice. For seven class periods before the incentive plan is introduced, the teacher records the number of pages of practice turned in daily by each student. The incentive plan is then introduced and that procedure is followed daily for a week, during which time the teacher also records the number of pages practiced each day. The incentive plan is then stopped, but the teacher continues to record the pages practiced for seven class days following the end of the incentive plan. In this way the teacher will see whether the amount of practice done during the incentive plan was greater than that done before the plan was introduced and also whether the effect of the plan carried over afterwards.

This design, known as the **Time-Series Design** (Campbell and Stanley, 1966), is depicted this way: $O_{1a}\ O_{2a}\ O_{3a} \ldots O_{7a}\ (T_a\ O_{8a}\ O_{09a} \ldots O_{14a})\ O_{15a}\ O_{16a} \ldots O_{21a}$. This says that seven observations were made prior to the introduction of the treatment, during which time seven more observations were made (observations 8 through 14) and at the end of the treatment for seven days, seven more observations were made.

This procedure does allow some comparison of what happens before, during, and after treatment, but in order that the influence of the incentive plan may be assessed after it has been stopped, the teacher must be sure that nothing else is introduced that could influence the after-treatment results. Suppose, for example, that on the fifth day of the incentive plan, the teacher suggests to several members of the class that they listen to tapes of the material they are to practice while they are copying it from their books. The results may appear to show that the incentive plan had some carry-over effect when in reality listening to tapes may be what accounts for the after-treatment amount of practice.

Although this design overcomes some of the limitations of the two previous designs, the statistical analysis is very complex and usually difficult to interpret.

Static-Groups Design

Let us now recall the business mathematics teacher in the first illustration, the one-shot case study, and compare that with the following procedure. The teacher introduced a different way of doing addition problems and then measured the skill with which students could work addition problems. Let's say the teacher has two sections of business math and introduces the new procedure in one section but not in the other. At the end of the experimental time, an addition test is given to the two groups, the "experimental" group having the higher scores.

Such an experiment may be represented in symbols this way:

$$\begin{array}{c} T_a O_{1a} \\ \hdashline O_{1b} \end{array}$$

This says that Treatment A was used with Group A but no treatment was introduced to Group B. After the treatment was over, an observation was made of the performance of both Groups A and B. The dotted line tells us that the two groups were not equated at the beginning of the experiment.

This is called a **Static-Group Design**, which Campbell and Stanley (1966) describe as a pre-experimental design.

Even if statistical tests indicate that the difference in test scores between the two groups was not a chance occurrence, what do we really know about the effects of the experimental treatments? How do we know that the two groups were equal in arithmetic ability at the start of the experiment? Perhaps the experimental group was composed of young people who had more ability than the group using the "traditional" method. Perhaps the teacher liked the new way of doing the problems and thus motivated the experimental group more than the other group. If the experiment was carried on over quite a length of time, some of the students may have dropped out. Since we don't know how similar the groups were at the beginning of the experiment, we do not know the relative effect of dropouts on the structure of the final group on whom post-experimental observations were made. Thus, we can see that just having two groups to compare is not a great improvement over the one-shot one-group design. We still cannot say that the findings would apply to any groups other than the ones involved in the experiment.

Posttest Design with Randomization

In the experiment involving drills in typing, the teacher could supply students within the same class with different types of drill copy, A and B, with appropriate instructions for its use recorded on tapes for students to listen to over earphones. In such a situation, a major decision to be made is "Who should use which drill materials?" The answer lies in this design:

$$R\;T_a\,O_{1a}$$

$$R\;T_b\,O_{1b}$$

The **R** indicates that the students are assigned at random to the groups to receive each of the treatments. The treatments (T_a and T_b) are carried out, after which the observations are made; that is, an observation (measurement) is made of the straight-copy skill of both groups A and B.

Since the students had received the same treatment up to the time of the experiment, their prior experience should not have an effect on the experiment. The teacher need not test for typing ability prior to beginning the experimental treatment. But other things, native ability, motivation, attendance, as well as different degrees of typing skill, could in some way influence the performance on the criterion test at the end of the experiment. These effects, some of which we cannot or do not wish to measure, are best equated in the two groups by randomly assigning the students to the treatment groups. The presence in varying degrees of abilities, motivation, etc. of the students in the two groups is likely to be counterbalanced by randomization. The larger the groups one randomizes, the greater is the likelihood that they will be balanced groups. If pretesting is not important or if pretests may influence the results, the **Randomized Posttest Control Group** is suitable. It is a true experimental design, since **randomization** allows us to make statements with a specified degree of confidence that any differences may or may not be attributed to the experimental variable.

The second group of students in this experiment is not, strictly speaking, a control group—it is a second treatment group. A control group is one that receives no treatment at all. In most educational research, we are interested in comparing two or more instructional methods or sets of teaching materials, or other variables, rather than comparing a treatment with no treatment at all. One such study, however, in business education in which a true control group is appropriate might be a study of change in attitudes toward certain consumer economics topics, when a treatment group may be compared with a group of students who have not been enrolled in any business or consumer economics classes. Or we might have three groups: Group A is assigned reading from textbooks, Group B reads from popular magazines and newspapers, and Group C is a control group that receives no formal classroom instruction in consumer economics.

Pretest-Posttest Design with Randomization

If, in such a consumer economics experiment, we felt that students' reading ability might affect success in reading the various materials (and thus be directly related to the achievement test scores), we might want to pretest the students in the two groups to see if they have equal reading abilities at the start. This procedure may be represented this way:

$$R\ O_{1a}\ T_a\ O_{2a}$$

$$R\ O_{1b}\ T_b\ O_{2b}$$

This says that students were randomly assigned to Groups A and B. A pretest (O_1), a reading ability test, was given to both groups. The treatments were then introduced to each of the groups (the two groups were formed within the same class); and following the treatment, another test, one of achievement (O_2), was given to each of the groups. Both groups were treated alike except for the experimental treatment itself (different types of reading materials).

In this design we know whether or not the groups are equal initially in reading ability. If the two groups do not have equal ability in spite of randomization, certain statistical procedures may be carried out on the criterion measures (O_2) that will in effect remove from the criterion scores that part attributable to initial reading ability.* This is an example of statistical control. Through randomization, we also control any other characteristics of the students that might influence the results of the experiment.

This design controls for "history," occurrences in the situation that could affect a group's performance. If we can randomize within the same

---

*For a brief but lucid explanation of this technique, analysis of covariance, and its limitations, see Kennedy, J. J. The use and misuse of analysis of covariance. *NABTE Review*, 1977, **4**, 10-11.

class, any extraneous event would affect both groups in the same way; for instance, if the teacher was absent for three days and the class was taught by a substitute teacher, or if several fire trucks, sirens wailing, pass the classroom during the testing period.

The Four-Group Design

The Pretest-Posttest with Randomization Design, however, does not let us predict what the effect of the experimental treatment would be in groups that had not been pretested. This might be of consequence in a study of attitudes in consumer economics, where a pretest might affect the motivation and attitudes of students in such classes. In order to determine whether the pretest itself has an effect on experimental outcomes, we might set up a Four-Group Design (Campbell and Stanley, 1966) such as this:

$$\begin{array}{llll} R & O_{1a} & T_{Aa} & O_{2a} \\ R & O_{1b} & T_{Bb} & O_{2b} \\ R & & T_{Ac} & O_{2c} \\ R & & T_{BD} & O_{2d} \end{array}$$

Let's apply this design to the situation in which the teacher wishes to determine the effect of two instructional approaches on attitudes toward selected consumer problems. In order to know whether or not the experimental variable will affect (change) attitudes toward certain consumer problems, the teacher will need to know what the students' attitudes are before the treatments begin. An attitude scale administered before an experiment may influence the outcome of the study; so to determine the extent of this effect, we first assign students randomly to four groups, **a**, **b**, c, and **d**. A pretest (O_1), an attitude scale, is given to two groups, **a** and **b**, selected randomly from the four groups but is not administered in groups c and **d**. Treatment A is then randomly assigned to two groups, in this case **a** and c; that is, to one group that receives the pretest and one group that does not. Treatment B is then assigned to the other groups. Following the experimental period, a posttest, O_2 (a second administration of the attitude scale), is given to all four groups. Not only are we now able to compare the attitudes of the two groups exposed to the two treatments, but we can also compare attitudes of students who did and did not receive the pretest.

Also, by randomly assigning students to the main treatment groups, we have attempted to equate the groups on other factors, perhaps unknown, but they may be such factors as socio-economic level, general academic ability, etc., that may influence the outcomes of the experiment.

Many variations of these designs may be planned, also. We could compare the scores made by boys and those made by girls. We could also divide the treatment groups into subgroups according to their general abilities; we could perhaps get data regarding their socio-economic status and compare students in the different classifications. This process is known as blocking. We can also study many combinations of all these factors, although whenever a factor is added to a study, a greater number of subjects is required.

Randomization may also present some problems when the students are to be blocked, because if subjects are randomized before they are blocked we may find that we do not have enough students in one or more of the subgroups (sometimes referred to as "cells"). Sometimes one has to set the boundaries of the categories into which the subjects are classified rather arbitrarily and artificially in order to assure having sufficient numbers of subjects in the cells.

Above all, however, one must plan all comparisons one wishes to make BEFORE beginning to gather any data. Knowing something about possible analyses of the data influences the questions (and hypotheses) one asks in a study; and, on the other hand, the questions one asks governs the design one might use in order that appropriate statistical analyses may be made of the data to get the answers to the questions. The processes of formulating the problem and planning the analyses are inseparable.

Also, the more elements one introduces into the research design, the more complex the statistical analysis becomes and the more difficult the interpretation of the results becomes. Such complex designs are best left to those individuals who have a sound statistical background or who are willing to gain an understanding of appropriate analyses. These designs are included here to illustrate that the more carefully a study is designed beforehand, the better the variables may be controlled so that the conclusion that the procedure under study does or does not make a difference in the outcomes may be accepted with a specified degree of confidence.

Notice that in most of the designs discussed so far, randomization of the subjects to the treatments is an essential part of the design. With a large enough number of experimental subjects (and this **does not** mean thousands—or even hundreds), randomization permits broader generalizations, since the bias in the sample is reduced. Randomization should reduce the probability of any systematic characteristic appearing in one of the groups; that is, the likelihood is reduced that a disproportionate number of the brightest (or poorest) students will be in one group or more students from high socio-economic levels, etc.*

---

*Even with randomization, the unlikely can happen. Using a formula for measuring the difficulty of dictation materials taken in Gregg shorthand, an experiment was conducted to see if students writing Pitman shorthand would make a different number of errors on letters classified as hard, medium, or easy according to the formula. The results were as expected, with a greater number of errors on hard than on medium difficulty letters, and the fewest on the easy ones—except in one class, where the results were exactly opposite. Students had been randomly assigned to groups taking the various letters. In examining the characteristics of students in that class, the investigator found that **all** the best students in the class were in the group who wrote the hard letters and all the poorest students had been randomly assigned to the group who wrote the easy letters. Small numbers of students were involved.

Non-equivalent Control-Group Design

Some experimental procedures, however, are such that only one of them may be used in any one class. Also, most school systems may be unable or be unwilling to assign students randomly to particular class sections. This is a common situation in educational research. Campbell and Stanley (1966) propose for these situations what they classify as quasi-experimental studies, in which intact classes are used. Such a design may look like this:

$$\begin{array}{c} O_{1a}\ T_A\ O_{2a} \\ \hdashline O_{1b}\ T_B\ O_{2b} \end{array}$$

The label attached to this design is **Non-equivalent Control-Group Design.** The dotted line shows that the two groups being studied are taken as is; the students were not randomized individually to the treatments. Measures are taken on some variable related to the criterion measure (perhaps a measure of prior knowledge of the content being learned) prior to initiating the experimental procedure (O_1), and the groups are compared on those scores. The experimental procedures (T) are carried out, following which criterion measures are taken (O_2). If the groups did not differ on the pretest score, the researcher proceeds with the analysis of the O_2 data without having to account for any initial differences of the groups. However, if the groups should differ initially, a more elaborate statistical analysis may be called for that removes from O_2 scores that portion that is contributed by the factor measured by the pretest. According to Tuckman (1972, p. 118), this is an essential step in this design, although Kennedy (1977) disagrees with this use of analysis of covariance.

Two other important considerations regarding this design are often overlooked in the research reported in business education and must be pointed out.

1. If intact groups are used as the subjects in experiments of this design, the analysis must be done on group means and not on individual scores. Analysis of individual scores in this case is an incorrect procedure. Thus, in order to have a meaningful comparison, the experiment should be conducted in several (12 to 15 or more) pairs of classes. Hence, we have a larger sample of scores to analyze rather than a sample of one for each treatment, which we would have if we carried out the experiment in only one pair of classes. The greater the number of scores we can use in our analysis, the more reliable the outcomes are likely to be. A more detailed explanation of this statement will be made in the "Analysis" section of this chapter.

2. The intact classes selected for study should be as much alike as possible and generalization made only to the schools involved in the experiment. The readers are then left to judge whether or not their own classes are like the ones that supplied the data for the experiment, and they can decide whether the conclusions will apply to their particular situations. In any intact-group design, then, analysis of individual scores is invalid

procedure, and generalizations cannot be as broad as they are when individuals are randomized to the treatment groups.

Controlling Variables

The pre-experimental observations and randomization just discussed insure that, as much as possible, any differences between groups may be attributed to the experimental method under study. The precautions taken, the procedures used for accounting for elements other than the experimental variable, are the **controls** of the experiment. In general, they affect the validity of the experiment, insuring that the findings are due to the experimental variables and not to something else and also insuring that the findings would obtain in other similar situations.

Control procedures include such things as pretesting, equating groups, using valid and reliable measures, controlling the time spent by both groups on the learning procedure, and other related matters, such as teacher effect (by having each teacher teach both methods, assuming, of course, that the teacher is equally proficient in and enthusiastic about each of the methods). Teacher effect is also another reason for using as many pairs of classes as possible; then whatever prejudices teachers may have toward the treatments are more likely to be counterbalanced in a greater number of experimental groups. Carefully directing and checking on the exact procedures used in the groups to assure that all participants are treated the same way and assuring that testing procedures are identical in all groups are also important elements in controlling experiments.

Generally speaking, three basic types of control may be employed: selective manipulation, physical or mechanical manipulation, or statistical control. Selective manipulation has to do with the use of appropriate experimental subjects, seeing that they are the proper types of students, for instance, or teachers or equipment for conducting the experiment. One should know, for example, that the teachers in an experiment are prepared and able to teach the subject matter in the classes under study, that the students are capable of handling whatever is being taught.

Physical manipulation or control includes such procedures as providing explicit detailed directions to the participants about the conduct of the experiment, keeping practice materials the same, or the amount of practice the same (unless, of course, the length of practice period is the variable being studied), and supervising the participants during the experimental period. We can also use mechanical means in such things as presenting copy in typing or shorthand classes by means of a tachistoscope, or perhaps administering shorthand dictation tests via tapes.

Statistical controls are those in which that amount of the criterion variable is removed that may be attributed to the factors measured in the pretest(s), or by which the interaction of the variables may be measured. Randomization is perhaps the most important element in statistical control.

Internal and External Validity

Any researcher, then, wants to be as sure as possible that observed differences may indeed be attributable to the experimental variable; he is concerned, in other words, with **internal validity** of the experiment. He must also be concerned that the findings will apply to other similar situations; that is, with **external validity**.

Several authors of research methodology books (Tuckman, 1972; Borg and Gall, 1971; Campbell and Stanley, 1966) identify factors that must be controlled to insure internal validity. These factors are listed here, along with examples of procedures that may be used to control them.

1. **History**. History refers to some unplanned event occurring during the experiment that could affect the outcome; for instance, the electricity going off during a test requiring use of office machines. Randomizing students to the different experimental methods within a single class controls history in that any unexpected event would affect both groups similarly. However, if two intact classes are being compared and this event occurs in only one class, we would not know what the effect would be of having to retake all or part of the criterion test because of some difficulty during its administration.

2. **Maturation.** If the experiment continues over quite a long period of time (several weeks or months) or if the data-gathering period is lengthy, the fact that the students were in the class that long may in itself account for some of the change that takes place instead of the method being studied. If over that time several measures were to be taken, but memory effects and skill being developed could affect the scores, the instruments would be administered in different orders in each of the data-gathering sites.

3. **Testing.** This is the effect of one test on another, as the effect of administering an attitude scale in a consumer economics experiment on the attitudes held by the students at the end of the experiment. If this is likely to happen, no pretest should be given, but students should be randomized to the treatments to control for pre-experimental differences in the groups.

4. **Instrumentation**. Instrumentation has to do with the way the data are gathered and the nature of the instruments used to supply the data. The validity and reliability of the measures must be as high as possible. In order for the results to apply in another situation, one must know that all the measures are reliable, that the same results would obtain if a test or scale was used again in a similar situation. Unless the criterion measure (and any other measure) is demonstrated to be valid for the particular group in the particular experimental situation, the research effort will have gone for naught.

For example, in one study, the researcher compared the ability of two classes to produce mailable letters as a result of certain teaching procedures. In one group, emphasis was placed on teaching punctuation and spelling skills in individual sentences rather than in sentences in letters. All intermediate tests were on individual sentences rather than in sentences in letters.

In the other group, no practice was given on individual sentences, and all punctuation situations were presented and practiced in letters. All intermediate tests were mailable letters. The criterion test for both groups consisted of three letters evaluated on all elements constituting mailability, including placement and letter style. Since the first group had not practiced transcribing letters, a mailable letter test was not a valid test for them, even though the test covered the punctuation and spelling on which they had been drilled. It would not be an appropriate basis for testing the difference between the two groups.

Besides the validity and reliability of the tests, the testing conditions must be the same for all experimental groups. If the scores on an achievement test are the criterion score, one would not just supply one copy of the test for each group and ask the participating teachers to duplicate the test. Some might do a professional offset job, others might use a stencil duplicator, and a few might quickly make a heat transfer spirit master and run the required number of copies. The instrumentation in the groups, then, is different since the print quality and readability of the tests might be different in the different groups.

5. **Selection.** Selection of subjects is of great importance. We have seen the importance of randomization in eliminating bias and in controlling unknown sources of contamination of the results. Some people, however, may feel that matching procedures should be carried out instead. Matching of subjects, as a control, poses several problems. Exact matching is usually impossible, so the resulting groups are not as alike as one might desire them to be; thus, an unmeasurable bias is introduced. If one tries to match on more than one variable, one finds that several individuals cannot be matched on the second variable once they have been matched on the first. Thus, one loses a great many possible subjects, and the very fact that they are eliminated biases the remaining sample. The elements on which the subjects were not matched could be the source of considerable bias.

Statistical control procedures allow all available subjects to be used. Over 300 students were available for participation in an experiment reported a few years ago. By the time the researcher had "matched" students on four variables, the final sample was constituted of just 19 pairs of students. This is an expensive, wasteful procedure. However, if close matching on one variable is possible, then assigning each individual of each pair to the treatments (at random) is the procedure of choice.

And, of course, the subjects for study must be representative of the group to whom the results will be generalized. One doesn't carry out an experiment or survey on college freshmen and expect the results to apply to tenth graders; one would not study university freshmen and generalize to all first-year post-high school students.

6. **Statistical regression.** A common phenomenon in the classroom is that the students who score lowest on a test at the beginning of an instructional procedure are likely to gain more during the instructional period than do

the students who score the highest on the pretest. That is, the scores of both extreme groups will not deviate as much from the criterion test mean as their pretest scores deviated from the pretest mean. This phenomenon is known as **regression effect**. A group with widely ranging pretest scores will seemingly gain less than a group whose scores all cluster around the initial mean. This may disguise the real effect of the experimental procedures. Thus, if two such groups were compared, the results may really be accounted for by the nature of the group rather than by the experimental treatment. Experimental groups must be as much alike as possible at the start of the experiment.

If the groups do happen to be of different abilities or score differently on pretests, statistical control procedures are possible if the subjects are assigned randomly to the groups.

7. **Mortality.** What is the effect on the composition of the groups if some of the students drop out of the group before the termination of the experimental procedures or if some students are not present for the criterion testing? Unless we have some prior knowledge (pretest) of the "influencing" characteristic of the group, we cannot assess the effect of dropouts from the groups. Perhaps again we have a biased sample at the end of the experimental period so that we cannot know whether any observed differences may come from the nature of the group remaining or from the experimental procedures.

Suppose we have an experimental situation in which Group E and Group C had equal mean scholastic aptitude scores, with equal variance at the start of the experimental period. However, as the experiment progressed, a greater proportion of students dropped out of Group E than out of Group C, and those who dropped from Group E were those with the lowest scholastic aptitudes. Because we had determined the aptitude level before the experiment began, we can now determine to what extent the dropouts changed the aptitude level of the group taking the criterion test and can apply statistical controls if the two groups do differ as a result of dropouts.

8. **Interaction of variables.** If one experiments with two groups using two methods, but does not randomize Ss to the treatments nor take a pretest measure of student ability or previous achievement when it may be closely related to the criterion measure, actual differences may not show up since the differences may be masked. For instance, if no knowledge of initial ability of the subjects is available, the researcher cannot find out that an experimental method may actually produce high scores for students with a particular level of ability but not for the other levels; but the statistical tests overall may show no difference between the experimental groups as a whole. In an experimental design that does not account for initial ability and control for it (for instance, by blocking students on their ability scores), the interaction of the ability level with the method may be masked.

However, using a design in which students within each of the treatment groups is grouped according to levels of ability, for instance, is a means of utilizing this information as part of the experiment. This is a strong control that produces much valuable information.

Thus, controlling variables in experiments or other types of research may be accomplished in three ways: (1) physical (including mechanical) manipulation, (2) selective manipulation, and (3) statistical manipulation.

In physical manipulation the control is actually imposed in handling the situation—regulating the exposure time of symbols in presenting shorthand words by tachistoscope, presenting typing copy at a given rate on a controlled reader, administering shorthand tests via tape so that all subjects have the dictation given in the same way by the same dictator, or having all subjects in the same room with the same kind of lighting and other equipment.

Selective manipulation involves choosing appropriate subjects properly, as well as selecting appropriate materials and/or methods. Subjects are to be selected because they are representative of the population to whom one wishes to apply the results. Selection of materials may mean seeing that they are of equal difficulty (unless, of course, one is interested in determining effects of difficulty on learning) or equal in length, etc.

Statistical manipulation is the use of some initial measure to equate the groups rather than matching the individuals in the groups; it may involve taking account of interactions between variables and interrelationships among variables. The first procedure removes initial effect; that is, it removes the portion of the differences that may be attributed to the related variable. If we are studying the effect on achievement of two methods of teaching and we have reason to believe, for instance, that reading ability and general scholastic ability are also related to achievement, we would not have an accurate assessment of the outcomes of the experimental procedures if we neglect to determine that part of the outcomes that should be attributed to reading and general scholastic ability. Or, we should divide the experimental groups into subgroups according to appropriate initial ability measurements to see if the procedures effect outcomes for any one of the groups more than the others.

Analyzing Experimental Data

The data gathered in experimental research are a set of scores for each treatment group. As we know, a student's test score is actually made up of several factors other than just knowledge of the subject matter. We know that each student taking a test brings to it many other factors: some ability, anxiety in different degrees, a certain amount of knowledge, and motivation in varying amounts. Also involved in the test score is the nature of the test itself, its validity, its reliability, its difficulty, its sampling of the knowledges and skills that the student was to have acquired from the instruction in the class. Other things also may contribute to the scores students make: the physical make-up of the test, the atmosphere in which it is given, the temperature and ventilation in the room, etc. All these things added together make up the students' scores, which vary one from another.

Analysis of Variance

We know, also, that different groups of students as a whole perform differently on the same test. It is these variabilities in the scores, both of individuals within groups and of groups themselves, that form the basis of the analysis of data in experimental research. The data are usually summarized and presented in a table of Analysis of Variance, similar to Table 4-1, which you have seen in research reports. Table 4-1 shows an F value and some other numbers, all of which were calculated from the fictitious data given in Table 4-2. The F statistic tells us whether or not the differences found can, in fact, be attributed to the experimental variable. But before going into an explanation of what the numbers in such a table may mean, let us define and explain some of the terms in the headings.

Table 4-1. Analysis of Variance

Source of Variation	Sum of Squares	df	Mean Square	F
Between (treatments)	4.05	1	4.05	.30
Within (groups)	242.50	18	13.47	
Total	246.55	19		

Variance. The technical definition is this: Variance is the mean of the squared deviations of each of the scores from the group mean. But what does that say? After tests are scored, a common practice is to divided the total of the scores by the number of papers to determine the mean score for the class. Another piece of information about the test that teachers should know is, "How much do the scores vary, that is, deviate, from the mean?" If one tries to calculate the mean of the deviations of the scores from the mean, however, one runs into a difficulty. It is this: If the differences between the mean and each of the scores are totaled, taking into account the direction of the deviation (plus signs indicating scores above the mean and minus vaues for scores below the mean), the sum is 0. To avoid this problem, the differences are each squared, making all values positive, before they are totaled. In the Analysis of Variance table the total of the squared deviations appears in the column labeled **Sum of Squares** (SS).

Table 4-2. Scores of Two Treatment Groups.

Group A n = 10	Group B n = 10	Total n = 20
10	7	17
12	20	32
13	13	26
16	12	28
19	10	29
18	19	37
12	16	28
17	12	29
16	14	30
10	11	21
X = 143	134	277
$\overline{X}$ = 14.3	13.4	13.85

However, notice that three values are given in the SS column in Table 4-1. Three different sums of squared deviations have been calculated in this analysis of variance:

(1) The sum of squares of the deviations of each of the 20 scores from the grand mean (13.85), that is, the mean of all 20 scores. This SS is the "Total," 246.55 on Table 4-1.

(2) The SS of the deviations of the two group means around the grand mean. This is the "Between" value, 4.05. Referring to Table 4-2, you will see that the two means (14.3 and 13.4) differ from each other and they differ from the grand mean (13.85).

(3) The deviations of scores of the individuals within each group around their respective means. This is the value of 242.50, labeled "Within."

The calculation and use of these three variances should become clearer as you read further. As we know, we cannot control all the elements in an experiment that make up the total variance in the criterion score. To discover how much of the total variance comes from elements other than the experimental variables, we merely subtract the "Treatment" SS (4.05) from the "Total" SS (246.55), resulting in a "Within" SS of 242.50. This remainder is sometimes known as the **residual** variance, or, more often, the **error** variance. In this case, error does not refer to something that is incorrect; rather it refers to that part of the total variance for which we cannot specifically identify the source.

We have now identified the SS's, the sums of the squared deviations from the means. However, since variance is the **mean** of the squared deviations, we must divide these SS's by their respective degrees of freedom (**df**'s).

Degrees of Freedom (df). Ordinarily in calculating a mean, we divide the total of the scores by the number of scores involved. However, in calculating variances, for mathematical reasons beyond the scope of this bulletin, the divisor is the degrees of freedom, the 1 and 18 in the **df** column in Table 4-1. Here is a simple example to explain this term: If we wish to find the mean of the scores 10, 12, 16, 13, and 17, the sum of which is 68, we will divide the total of 68 by 5 to get the mean, 13.6. Let's assume we know four of the five scores that total 68; namely, 12, 16, 10, and 13. Then the fifth score **must** necessarily be 17, since no other value would result in a sum of 68, given those other four values. Four of the five scores that add to 68 can be any value; but in order for the five values to add to 68, the fifth one must necessarily be the total less the sum of the four known values.

Let's look at another example. Seven values add to 102. Six of those values could be any six numbers you think of; but because those six are what they are, the seventh one is fixed (total, 102, minus the sum of the other six values). In the first example, four values were "free" to vary, to be any value; in the second example six values were "free." In the first example, we have four degrees of freedom and in the second example, six. The degrees of freedom are **n-1**, which is one less than the size of the sample. Notice the lower case **n**, which customarily refers to **sample** size, while **N** indicates a population.

Looking at Table 4-2, we see that the mean of Group A is 14.3; Group B, 13.4; and the Grand Mean (mean of all 20 scores) is 13.85. In analysis of variance, we are testing to see if the two group means deviate from the grand mean more than the individual scores deviate from their group means. Thus, we in effect calculate the mean of the means of the two groups. In order for the mean value of the two means to be what it is, one of the group means being what it is determines what the second one is. Thus, we have one degree of freedom, since only two group means were involved in calculating that mean.

We also work with the grand mean, composed of 20 scores, of which, if 19 can be any value, the 20th is fixed, and we have 19 degrees of freedom. Then if we subtract the 1 df for the treatments from the 19 df's of the Total, we find 18 df's for the Within (Error) term.

Mean Square. This is actually the variance; that is, the SS divided by the df. The Mean Square (MS), then, is the amount of actual variance due to each of the breakdowns listed in the table—in this case Between (treatments) and Within (groups) or residual variance. (By the way, if you calculate the square root of the variance of each group, you would be calculating their respective standard deviations.)

The F Statistic. We said above that in analysis of variance, we are testing to see if the two group means deviate from the grand mean more than the individual scores deviate from their group means, and to see if this difference is greater than we would expect by chance. However, since deviation of scores from their respective means adds to 0, and the deviation of the two group means from the grand mean also adds to 0, we must use the squared deviations (variances) from the various means instead of the actual means themselves to test this hypothesis.

By using the Sum of Squares and the Mean Squares shown in Table 4-1, we have determined how much of the total variance may be attributed to each of the sources of variation. We still do not know, though, whether the amount contributed by the treatments actually differs more than expected by chance from the Error (Within) variance.

We test that difference by first calculating the ratio between the treatment variance and the within variance (mean squares), resulting in an F ratio. In this illustration we divided the Treatments MS (4.05) by the Within MS (13.47) to produce an F value of .30. This F value is compared with F-value standards given in F Tables in most statistics books.

Mathematicians have calculated how large the ratio should be, according to the degrees of freedom for the two variances being compared, in order that we may know that with either 95 percent probability or 99 percent probability we can attribute the difference to the experimental treatment. We compare our calculated F value (.30) with the F value given in the F Table for 1 and 18 df's. If our value is as large as or larger than the printed value, we can **reject** the hypothesis of **no difference** between the groups' performance. The F values given for 1 and 18 df's is 4.41; but our F value was .30, so we can say that the treatment did not contribute to differences between the groups.

Interpreting the Findings. What we are actually doing in an analysis of variance is testing the hypothesis of **no difference** between groups' performance on the criterion measure. This is the **null hypothesis.** The statement used in reporting such findings often reads this way: "The difference was significant at the 5 percent level." This means that we may reject the hypothesis of no difference in the performance of the two groups with 5 percent probability that we do so when, in fact, the hypothesis is true.

In testing hypotheses, the level of significance is set **before** the data are gathered. It is part of planning the study. It indicates really how large a risk one is willing to take in being in error. How "costly" would acceptance of a false hypothesis be, or the rejection of a true hypothesis? In experiments for developing a potentially dangerous drug, for example, the level of confidence would necessarily be set much higher than it need be in educational research. Would we be willing to take a risk of one in a hundred, one in a thousand, or one in 10,000 that a new drug will not produce a debilitating side effect, or would you want that probability to be even higher? However, in education, if our experimental method is sound pedagogically and is firmly based on sound theoretical bases, will the probability of rejecting a false hypothesis be of serious consequence? Thus, 5 percent probability of error or even a 10 percent level of significance may be acceptable.

In many studies done with great numbers of subjects, much is made of the fact that the observed differences are significant, yet the differences may be only 3 or 4 points on a hundred-item objective test. With large enough numbers of subjects, a difference of only 2 or 3 points may be significant; yet in a criterion test of 100 points, how much more does Student A, who has a score of 85, "know" than Student B, who scores 82? Perhaps if just a few other questions had been asked instead of some of those on the test, the two students' scores might have been reversed. Thus, one should not only be concerned with whether the difference is statistically significant, but should also pay attention to the actual size of the difference—is it a practical difference?

T-Tests. Along with F tests, one must mention t-tests. A t-test of significance may be used to test the hypothesis of no difference between two groups, while two or more groups may be involved in experiments using the F test. The t-test is interpreted in the same way as the F test. As more people are becoming acquainted with analysis of variance, F tests are gaining in use. They allow greater control of more variables, and they give information about the relative contribution of several variables, as well as about the interactions among the variables. The only thing that a t-test can tell us is that two groups do or do not differ in their mean score performance on the criterion measure. (The ANOVA problem used as the example had only two groups to keep the calculations as simple as possible.)

Options in Analysis of Variance. In the above discussion, we have used as an illustration a situation in which two treatments were compared in which the students had been assigned at random, without any pretest, to the groups. This basic analysis may also be used in more complex designs; for instance, in comparing the variances of subgroups within the main groups as well as the variances of the main groups themselves. However, then we have more than one source of variance besides the error variance. We could, for instance, depict the study of the effect of two types of reading assignments in basic business classes, dividing each treatment group into subgroups of high, medium, and low reading ability. This design may be shown this way:

Ability Level	Treatment	
	Textbook	Newspapers
High	X_{th}	X_{nh}
Medium	X_{tm}	X_{nm}
Low	X_{tl}	X_{nl}

In this study the total variance may be broken into three sources of variance besides the error variance, as in this analysis of variance (ANOVA) table:

Source of Variation	SS	df	MS	F
Treatment		1		
Level (of ability)		2		
Trtmt x Level interaction		2		
Within (Error)		(n-1) - 5		
Total		n-1		

Such an analysis is a two-way analysis of variance, since the data are classified in two ways: by two **treatments** and by the three **ability** levels. These are known as the **main effects**, while the third source of variation is the amount of the total variance that may be attributed to the **interaction** of the main-effect variables. A significant interaction tells us that the treatment affects one of the subgroups differently than it does another. If one finds a significant difference between the scores of the two treatment groups, one will have to consult the table of means and standard deviations of the groups that should accompany any presentation of analysis of variance to identify the group with the higher mean.

The data for our reading assignments experiment could also be further classified according to the sex of the students (provided this additional breakdown was planned **before** any data are collected!). The following table shows such a design:

Ability Level	Sex	Treatments	
		Newspapers	Textbooks
High	Boys Girls		
Middle	Boys Girls		
Low	Boys Girls		

Now we have three main effects: (1) Treatments (type of reading assignments), (2) Ability levels (high, medium, and low reading ability), and (3) Sex of the students. This is, then, a three-way analysis of variance, since the data are classified three different ways. These three main effects produce four interactions, as shown in this analysis of variance table:

Source of Variation	Sum of Squares	df	Mean Square	F
Treatment		1		
Level		2		
Sex		1		
Interactions				
T x L		2		
L x S		2		
S x T		1		
T x L x S		2		
Error (Within)		(n-1) - 11		
Total		n - 1		

F-values, then, are calculated for each of the three main effects and for each of the interactions. Each mean-square value is divided by the error mean-square to get the F value. The obtained F values are compared with the values given in an F table for the appropriate degrees of freedom to determine whether each main effect or interaction is significant. Significant differences between any of the main effects indicates, of course, that the means of the groups that were formed by the various classifications differ more from each other than would be expected by chance, that is, that the main effects account for more of the total variance than do other unknown factors (error).

Significant interactions may be even more important than the main effects from an educational point of view. In our illustration, for instance, significant interactions may give us information about the need for differentiating reading assignments for the different groups. These analysis of variance (ANOVA) designs testing interactions give us a great deal more information than a one-way ANOVA that tests only the main effects.

Actually in a three-way analysis of variance, we are testing seven different null hypotheses, which must be stated in the research report. Each research subproblem (question) must be stated in null hypothesis form if the plan of the study is to look for differences between groups. This is also true for all types of associational studies in which one is examining differences among groups, even though statistics other than F are used to test the null hypotheses posed by the researcher.

Other more complex designs may also be used, and newer techniques for analyzing the effects of variables on the total variance are feasible through greater accessibility of computers. Perhaps the most widely used of these is the multivariate regression analysis. In effect this tells us whether the total variance is reduced a significant amount (the percentage of reduction in variance is actually calculated) if a particular variable is eliminated from consideration. One of the main advantages of this technique is that randomization is not a requisite assumption, and non-quantitative data may be included as variables, such as prior study in bookkeeping. Details of these techniques are beyond the scope of this discussion; however, a detailed, lucid explanation of multivariate regression analysis is given by Houston and Teglovic in the 1971 NBEA Yearbook, Chapter 34.

Other statistics are also used for testing hypotheses, depending on the design of the study and the characteristics of the data being analyzed. One of the most useful and widely used of these is the χ^2 (Chi Square). It is used with data consisting only of frequencies rather than actual score data. The subjects under study may be classified on either quantitative or qualitative data. We could classify students according to level of ability (quantitative), according to whether or not they are employed part-time, or according to a teacher's assessment of whether they would be rated high, medium, or low on dependability or persistence or some other qualitative trait.

The figures in such tables represent the **number of students** in each category who are or are not employed or who rank high, medium, or low in specified traits; and the hypotheses tested might be that of no difference in employment status of students of different ability levels or of different trait qualities.

A common use of χ^2 is in analyzing opinions expressed on Likert scales. We could, for example, compare the opinions of student teachers and college student-teaching supervisors on an item like this:

> Student teachers should be allowed an option of whether they should receive Satisfactory-Unsatisfactory ratings or traditional letter grades.
> SA A ? D SD

The null hypothesis of no difference in the distribution of responses of college supervisors and of students would be tested. The χ^2 statistic is interpreted in the same way as the F test is. The obtained χ^2 value is compared with the value given in a Chi-Square Table for the appropriate degrees of freedom at either the five or one percent level of significance.

Chapter 5

Writing the Report

Since the ultimate purpose of most research in business education is to provide a basis for change or to evaluate the current status of the field, the research report must be written in a way that will encourage its being read. After one has spent the amount of time, money, and energy necessary for carrying out a respectable research study, the researcher cannot afford to produce a poorly written report of the study. Not only should the writing itself be good—grammatically correct, concise, and precise—it should also be well organized, presenting in clear, logical order the procedures, findings, conclusions, and recommendations. Enough detail must be included so that an individual interested in replicating the study could follow the report step by step and obtain similar results.

Certain information about the problem and details of procedure must be included and thoroughly explained. The study should be introduced by giving the background of the problem and information about how the investigator became interested in it, leading to a succinct statement of the problem. A foundation should be laid for the study in terms of similar research done prior to the study and also in terms of the theoretical base underlying the problem, as well as research that supports one's selected procedures. The need for the study must be established and presented along with a careful delineation of the purpose(s) of the study. Some of these topics are covered in a "review of the literature" section, but some of them form essential parts of the opening pages or are included in the procedures section.

The researcher outlines the procedures used in gathering the data, in controlling any variables that might influence the outcomes of the study other than the ones being studied, and in analyzing the data. The findings resulting from the analysis of the data then are presented, the study is summarized, conclusions are drawn from the findings, and recommendations are made.

Let us consider in more detail each of these parts of the research report and what is entailed in presenting each of them effectively.

Introducing the Study

The introduction, sometimes labeled "Background of the Study," serves just that purpose; the investigator explains how he became interested in the problem and why he felt the study was worth pursuing, establishes the need for the study, and sets forth its purposes. Researcher interest in the problem may stem from observations of the behavior of students under certain conditions. Or the researcher may be cognizant that current employment needs of businesses in the community seem to be different from those of a few years ago. This interest may actually be stated as a personal interest, but it is usually a general presentation of background reading or one's experiences with students or whatever led the researcher to the study being reported. Even though it is written in impersonal terms, it shows the researcher's interest in the problem and its significance to him and to business education. However one becomes interested, he explains the particular circumstances surrounding the problem, but without giving a complete history of the particular subject matter involved. For instance, if the researcher is experimenting with different teaching methods in bookkeeping, the teaching of bookkeeping need not be traced from the time of Paccioli.

Need for the Study

After giving the general background of the study, the researcher carefully establishes the need for the study, telling why the information resulting from the study is needed and showing the importance of the study. In many studies this will precede the statement of the problem, but in others it might more logically follow that statement. The need may be established on the basis of one's reading or on the basis of the findings of other studies; but however the need is developed, it must be supported with facts and figures whenever possible, giving all appropriate references. Often the need for the study is not labeled as a separate section of the first chapter and sometimes it is incorporated with the purpose. Davidson and Kilgore (1971) use the need for the study as the introduction to their journal article reporting the development of an objective test for measuring economic understandings of primary school children:

> The new materials in economic education emphasize the development of children's understanding of basic economic concepts or ideas, with these concepts to be taught as part of the structure of the discipline. In developing these materials two assumptions are made. First, it is assumed that an examination of the ideas or concepts contained in the materials will provide an interesting and meaningful experience for the children. Second, it is assumed that teachers will understand the concepts and present them in an effective manner.
>
> But the truth of these assumptions can only be tested by evaluation instruments to determine children's ability to learn as well as teachers' ability to teach these new materials—comparable to the Test of Economic Understanding at the high school level and the Test of Understanding in College Economics at the college level. Preferably, evaluation instruments providing benchmarks and indicating students' economic understandings should be available for use at the completion of both third and sixth grade. The following pages describe the efforts of the authors to develop a satisfactory testing instrument for use in the primary grades. (pp. 17-18)

Beyda (1967), too, incorporates the need for the study in one of the introductory paragraphs of her study of economic education programs in business firms. Although need is established, no specific section of the study has been labeled "Need for the Study."

> The need for economic education seems to be taken for granted, but the dilemma of what this education should consist and how and by whom it should be conducted still persists. Since both educators and businessmen are involved, a clarification of the responsibility of each group for providing economic education for the nation's citizens would be helpful. Such a clarification is dependent in part upon knowledge about the nature and extent of business activity in this area. Because of the lack of a comprehensive study in the last ten years regarding the economic education provided by business firms, no sound guidelines for future action exist. (p. 3)

Purpose of the Study and Statement of the Problem

The purpose of the study informs the reader about how the findings will benefit the field of business education. The purpose of the study is **not** the same as the statement of the problem. Rather, it spells out how the results of the study might be used and by whom. The purpose of a prediction study, for instance, is not to determine a set of predictive measures that will select those students most likely to succeed (or fail) in a course; that is a statement of the problem. The purpose is to supply information that will help counselors work with students in planning their programs of studies; to help shorthand teachers improve their courses by accomplishing more in the allotted time (since those students most likely to have trouble might be counseled into some other more profitable course for them); and to provide students with information about themselves so that if they show minimal potential for shorthand, they can, if they wish, elect some other business course that will better prepare them for the business world than would shorthand. The statement of the problem in the shorthand prediction study becomes then, the identification of those interests, abilities, and traits that are related to success in learning shorthand.

The following example, showing the distinction between the statement of the problem and the purpose of the study, is taken from Osborn's (1970) study of the relationship between shorthand reading and writing rates. Her statement of the problem is as follows:

> The problem of this study, then, was to determine the relationship between shorthand reading ability and the rate at which dictation could be recorded. (p. 7)

Osborn then expands upon the problem and gives several subproblems, each of which required a separate analysis of appropriate data, the answers to which, taken all together, answer or solve the main problem.

The purpose of this study was delineated by Osborn as being concerned with ways of teaching beginning shorthand that would result in better instruction and a saving of time, first stating the general purpose and then spelling out further implications of the purpose as follows:

> A related problem was to determine the relative effectiveness of three methods of teaching the reading of shorthand symbols in context and their effect on both reading and writing of beginning shorthand students. An attempt was made to answer the following specific questions:
>
> 1. Is speed of shorthand reading related to speed of taking shorthand from dictation?
> 2. What will be the effect on shorthand reading ability of beginning shorthand students if they are taught to read shorthand by silently reading the shorthand symbols in their textbooks while they listen to tape recordings of the reading materials?
> 3. Will scribble-writing while reading the material from the textbook and listening to that material from tape recordings have an effect on the ability of students to read from shorthand symbols?
> 4. Will a listening-while-reading approach to teaching have an effect on students' abilities to write shorthand symbols from dictation?
> 5. Will scribble-writing while reading the material from the textbook and listening to that material from tape recordings have an effect on students' abilities to write shorthand symbols from dictation? (p. 7)
>
> * * * * * *
>
> The purpose of this study was to provide information that might lead to a more efficient way to teach beginning shorthand. Because of the limited time available to develop shorthand skills in the modern curriculum, only those methods and techniques that are most effective should be used. If reading ability does influence the rate at which dictation can be recorded by the student, then reading skill should be stressed. However, if reading ability is not related to the rate at which the student can record dictation, the time spent developing reading skill might be better utilized in developing writing skill. By investigating three methods of teaching shorthand symbol reading...and also the effect of shorthand reading skill on ability to record dictation, a decision might be made as to whether or not reading skill should be stressed, and if so, how shorthand symbol reading might be taught more efficiently. (p. 6)

Madison (1969), too, clearly distinguished between the purpose of the study and the statement of the problem. She first made a general statement of the problem and followed it with more detail of just what the problem entailed:

> The problem is the identification of the need for employees and the general characteristics of office positions of technological nature.
>
> Specifically the problem is a survey* of the need employers in business and industry, and personnel directors in public administration, have for persons who would be employed in positions above the clerical or operational level and below the executive, policy-making level. Moreover, it is an investigation of the age, education, and experience qualifications for technological office positions as held by the employing persons for middle office positions. (p. 8)

---

*In the first paragraph Madison stated the problem as being one of identifying the needs; thus, the first sentence in this paragraph should read something like this:

"The identification of the needs for employees in office positions of technological nature and the general characteristics of those positions led to a survey of employers in business and industry, ..." (p. 4)

The **problem** is not a survey. The survey was the means for gathering the necessary data to identify the elements specified about technological office positions.

Further background of the problem was given, leading to a statement of the purpose of the study.

> This study is to provide information or facts to serve as a partial basis for the establishment of guidelines for the organization and development of college-level occupational curricula in business and economic education.... The study is to produce findings which may provide a partial basis for forming the objectives for the technological office curricula.
>
> A secondary purpose of the study is to provide guidance counselors, high school students, personnel directors, other employing personnel and office employees with information on the current and emerging office opportunities at the technological level and the general characteristics of these positions. (p. 8)

Theoretical Framework. The problem also should be based on a theoretical framework. In a study of the relative effectiveness of different methods of teaching, for example, the experimental procedure that is to be compared with some traditional procedure should be based firmly on theories and principles of learning. These bases should be spelled out very carefully in setting up the problem and establishing its significance. An excerpt from the discussion of the problem statement from Hayes' (1971) study illustrates how the theory underlying the problem may be presented and tied into the study.

> If the conditions of learning are understood, the learning situation can be managed or arranged so that information will be generalizable or transferable. Ten instructional processes that relate to purposeful learning are given by Klausmeier: (1) goal setting and clarifying objectives, (2) motivation..., (6) managing practice effectively, (7) providing for individual differences, (8) evaluating student performance...
>
> Usually students who enroll in shorthand have a general goal in mind, but the teacher may have to help the student clarify further the specific objectives and goals of the course.
>
> Motivation may be high at first while the learning is rather novel; but it may decline... Gagne reminds us that "the establishment of achievement motivation needs to be pursued vigorously and with great frequency." Variety of procedures and materials may help in keeping motivation high...
>
> Since this study is an attempt to find an effective method of practice, the process of managing practice effectively comes under study. Merely "repeating the same activity" does not produce the goals desired in the psychomotor and cognitive domains. Effective practice usually requires both productive thinking and physical activity, according to Klausmeier. One of the aims of this study is to determine whether a different method of practice will achieve "productive thinking."
>
> Students tend to "repeat the same activity" by copying shorthand notes from [illegible] textbook without any productive thinking; readiness to perceive, or attention, is frequently lacking. Therefore, the plan is to have part of the practice of the experimental group done from printed material... (pp. 10-12)

Delimitations and Limitations

An adequate statement of the problem also defines it very carefully in terms of its scope. The writer must set forth exactly the bounds of the **topic** being studied. These are the **delimitations.**

Delimitations are not concerned with the sample size or with definitions of a finite population; they are concerned with those parts of the **topic** or **problem** that normally might be considered a part of such a study but which, because of limitations of time, money, physical capacity, or other reasons, the researcher cannot or does not wish to include. For instance, in a study concerned with achievement in first-semester shorthand classes, mailable letter production usually would not be considered, nor would mailable letter format likely be part of the study. In other studies, the correctness of the students' shorthand outlines might not be a part of the study. These are delimitations. If in a study of consumer economics only that portion concerned with personal finance covering insurance, investments, and budgeting was under investigation, all other topics normally studied in such a unit would be delimitations of the study.

The delimitations "build a fence" around the **topic** under study, not around the population or the sample from whom the data are gathered. The selection of the population or sample is part of the design of the study and should be considered as part of the procedure. The dissertation by Ganser (1969) provides an example of appropriate delimitations:

> The study in no way attempted to determine the level of understanding economic concepts desirable for social studies and business education secondary education students. It attempted only to identify the extent of understanding economic concepts within the economic areas considered essential by the National Task Force.
>
> The study likewise did not attempt to determine the areas of economics which should be included for understanding. Since the Test of Economic Understanding was used to measure the extent of economic understanding for the subjects involved in the study, the basis for evaluation is therefore limited to the items within the test....
>
> The study did not in any way consider the teaching ability of the social studies or business education secondary education students. Rather, it considered the extent of understanding economic concepts. It was assumed that with a good understanding of the concepts, the graduates would be equipped better to become quality teachers of economic concepts than those with a poor understanding of the concepts. (p. 12)

Although the delimitations are those restrictions deliberately imposed on the study by the investigator himself, other facets of the study, by the nature of the research situation itself, may limit the researcher in making generalizations. These are the **limitations.** For instance, common limitations of questionnaire studies, especially if they are concerned with attitudes and opinions regarding emotionally toned topics, may be the instrument itself and the respondents' truthfulness in answering the questions.

Beyda's (1967) study provides an example of the difference between delimitations and limitations and also of how the data collected may not be complete since the population from whom data were obtained was from a restricted list. Also limitations are pointed out about the nature of some of the data since certain information to a certain extent had to be interpreted subjectively.

> Delimitations
>
> The study did not include programs of which economic information may be a part but for which it was not a primary objective. Excluded, therefore, were tuition-aid plans and scholarship plans which might incidentally provide some economic information. Analysis of posters, information racks, bulletin boards, and similar media was not included in the study.
>
> Excluded from the study was any evaluation of the programs or testing of the participants except that which was done by the firms who conducted the programs. There was no intention of categorizing the programs sponsored by business firms as providing accurate or inaccurate information, complete or partial information; being biased or unbiased, propaganda or education. (p. 5)
>
> The study is not intended to result in specific guidelines. It should be a help, however, in providing the basis of future development of such guidelines.
>
> . . . . . .
>
> Limitations
>
> This study is limited to the extent that the data obtained may not be the complete record of economic education programs particularly with regard to that which is historical in nature.
>
> The population for the study is the 1962 *Fortune Directory* of firms in the categories of industrials, insurance companies, commercial banks, merchandising firms, transportation companies, and utilities. They may not be completely representative of all business firms who have economic education programs since firms not listed by *Fortune* are excluded from the study. However, the programs sponsored by the firms listed will no doubt encompass the types most commonly conducted and will involve the greatest number of employees.
>
> The study is also limited to the extent that respondents may not be representative of the total population.
>
> Where qualitative data are a major portion of the information collected, a certain amount of subjectivity in analyzing and making deductions will no doubt be evident. All attempts, however, have been made to keep the degree of subjectivity to a minimum. (p. 5)

Another common limitation of survey descriptive studies that must be included in the research report (if it applies) is that imposed by a convenience sample rather than a random sample. A convenience sample restricts generalization to only the individuals responding to the survey. The researcher has no indication of the amount of bias introduced into the findings by a convenience sample, nor can he calculate the degree of reliability of the responses or the precision with which he may make any statements regarding the sample. (Helmstadter, 1970, pp. 327-28)

In an experimental study, a limitation is imposed by the situation in which intact classes must be used and for some reason treatments cannot be randomly assigned to the classes. The lack of randomization is the limitation. Another limitation in an experimental study might be something in the experimental situation that is so out of the ordinary that it may affect the outcomes more than would be desirable, as pointed out by Karaim (1968) in his study of the relationship between knowledge of theory and speed in writing shorthand.

> Another factor that should be accepted as a limitation of the study was that the subjects were exposed to an unnatural dictation situation while being photographed. Certainly, the presence of the camera and lights could not be considered part of a normal classroom atmosphere and may have affected the students' performances. (p. 12)

Thus, delimitations are those elements of the topic under study deliberately eliminated from consideration (always with good reason), whereas limitations are those factors inherent in the research situation that might affect the results but which the researcher must accept and recognize.

Definition of Terms

The definition of terms, too, is an essential part of presenting the problem. Those terms should be explained that have a unique use in the study, that are subject to several interpretations by the different readers of the study, or that are technical in nature. "Consumer finance" may mean buymanship to some individuals, personal finance to others, and money management to still others. A term such as "secondary school" also needs definition, for in some school districts this type of school includes the ninth grade while in other districts it includes only the tenth through the twelfth grades.

If any term is used in a way unique to the study being reported, the researcher may formulate a definition that fits that particular purpose. On the other hand, if a generally accepted definition is suitable, such as one given in a specialized dictionary, the writer may wish to use it. The source of the definition, of course, must be given. Other specialized or technical terms that might be unfamiliar to persons in other fields of education should also be defined, such as **standard words** in shorthand or **basic business**. The definition of terms is a most important facet of the research report, since without a common understanding of essential terms, the value of the research may be lost through misinterpretation.

An adequate introduction to a research report, then, will set the framework of the study, give a concise and complete statement of the problem and its scope, explain the purpose of the study and its importance, and carefully and precisely define any terms subject to different interpretations. This information, well presented, should lead logically to a review of related literature.

Review of Related Literature

The review of the literature, as indicated in Chapter 2, usually serves two purposes in the research report: (1) to set the theoretical base for the research and (2) to set the current research into perspective to show "the state of the art."

In setting the theoretical base, the writer may have read and reviewed writings and research in psychology, sociology, history, or any other field that forms a base on which his own research may be founded. Obviously theories and principles of learning form the basis for research in methods of teaching. In some studies, a review of the literature setting the theoretical framework might be handled more logically in connection with the statement of the problem than in a separate "review of the literature" section.

The second type of literature review, which may or may not be extensive enough to warrant a special chapter, shows how the present research fits into the whole scheme of things. It surveys the research previously done on the problem and evaluates what this research has and has not accomplished in solving the problem currently under study. Thus, in this review, the researcher should point out very carefully the similarities and, more important, the differences between that research and the current study. In order to do this effectively, however, the researcher must do more than consider the findings as they are reported. The findings must be examined critically in light of the research methodology, the specific procedures employed, the controls, the sampling procedure, and the measuring instruments used. In showing how previous studies differ from the present research, unclear or dubious procedures must be questioned, as illustrated by excerpts from two studies by Brandmeir (1969) and by Armstrong (1968).

The following excerpt from Brandmeir's study of the attitudes of business administration majors about business system environmental concepts illustrates how to question a procedural point in reviewing a study:

> ...this concerned an empirical study of a graduate school business management course as a socializing institution. Socialization was judged as the extent to which students' attitudes changed over time to be more like those of the faculty.
>
> Two student groups, business executives, and the faculty were tested on a before and after design using one hundred belief or value statements. A Likert-type scale having positions for strong and mild agreement versus mild and strong disagreement, but no neutral or undecided position, was used with a random reversing of scales. In less skilled test administration hands, the composition of the hundred statements of this design might have influenced the outcomes. The lack of a neutral position might also be subject to debate. (p. 35)

Armstrong, in reviewing the literature related to her study of different approaches to practice in developing skill in production typewriting, questions the means of reporting criterion scores; in discussing another study regarding the relationship between production and straight-copy rates, Armstrong (1968) appropriately raises questions concerning the reported findings:

The reader must consider three factors when comparing these figures:

1. The production rates included all the time required from the insertion of the paper to its removal. Thus, such important production non-typing activities as planning, machine adjustment, and proofreading were included in these rates.
2. The straight-copy rates were those given by the student himself in answer to the question on the data sheet, "What is your average per minute rate on 5-minute timed writings?" One cannot be sure whether or not the students replied in terms of gross words per minute, net words per minute, correct words per minute, or some other measure. Also one cannot be sure just how accurate the "average" was.
3. As nearly as this writer could determine, the production rate was found by taking the total number of words typed during the production test and dividing it by the number of minutes. No mention was made in the study of excluding work that was unacceptable or on which uncorrected errors had been found by the researcher.

* * * * * *

(The) students also had the opportunity to plan and arrange their problems, as well as adjust the machines, before the timing began. This, too, greatly influenced the problem rates obtained in the study. Moreover, his conclusions regarding the difficulty of the various types of problems may be valid only for the particular problems used in this study. For example, the typing of stencils had the highest percentage of transfer while typing a manuscript without footnotes had the fourth highest. However, the material to be typed on a stencil was ordinary straight copy and the only machine manipulations required were stroking, returning the carriage, and indenting the first line of each paragraph. The manuscript without footnotes was from typed copy, but it was mostly enumerated paragraphs which required the indention of all lines following the numbers. Since this type of indention is just the reverse of what students normally encounter in straight-copy typing, manuscript typing, and business letters, the difficulty of the problem seemed to stem from the nature of this specific material to be copied rather than from the type of problem.

Such a report of the literature as described here means that it is not an annotated bibliography. It is more than a series of synopses of studies or a series of quotations that merely abstract the procedures and findings of previous research. Good (1966) makes several suggestions to the writer of a research report for handling the review of literature. He says that summarizing research requires a balance between tedious detail and presenting only a superficial sketch of the research. This problem can be partially solved according to Good by grouping similar studies and giving detailed descriptions of a representative study or two. Good says, too, that weaknesses of unsound studies and the strong points of good studies should be indicated "in a constructive manner, but without ordinary adjectives of denunciation or praise as such" (p. 122). Good points out also that the review of related research is not "an appendage of a loosely related supplement" (p. 122), but that it is an integral part of the report and that "the purpose of the review is not primarily to produce a relatively complete annotated bibliography" (p. 122).

Scott and Wertheimer (1962) also recommend trying to formulate a single sentence or two that will summarize several studies on a particular subtopic in the area. By doing such summarizing, according to Scott and Wertheimer, the major results of research in a field may require no more than three or four pages and that "the more condensed and brief the summary, the more perspective it provides and the more apparent the inconsistencies become" (p. 33).

A topical organization lends itself to a better analysis of the previous research than does treating each study individually, either in chronological or alphabetical order. A topical consideration of previous research makes easier the comparisons and contrasts among several studies. It makes clearer the place of the present study in the overall research effort, as Brandmeir (1969) has done in the following paragraphs in which he pulls together several studies, making comparisons and contrasts:

> If one of the functions of a faculty is to influence the formation and change of attitudes, then attitude formation and change may legitimately be considered an element of student achievement. Note the contrast with the Astin report in which student achievement was concluded as being almost entirely a function of precollege characteristics. Faculty success in the Schein study is equated to student attitude formation and change and to student achievement. Maybe achievement is the amount of change in an individual's skills, perceptions, and attitudes, in conjunction with a satisfying self-perception rather than the attainment of an institutional norm. (p. 37)
>
> . . . . . .
>
> Useful as a frame of reference, Trow's report gives rise to expectation of considerable breadth of distribution in the responses in measurements of student attitudes with perhaps more than one mode. Business administration has often been characterized as entirely vocational. If this be so, response distributions should be quite narrow. If Schein's work on management education and McGinley's discourse on business school objectives are of consequence, then business student attitudinal measurements will show relatively broad distribution with some tendency to modes on specific attitude objectives, if not in the overall measures.
>
> A Michigan State University dissertation verifies Trow's report. It constitutes a significant support here because it is based on empirical data hypothesis testing, is only five years old, and is focused upon change rather than classification alone. Hodgkins dealt realistically with the dynamics problem by theorizing that the academic, vocational, and social goals of the college are shared by all the students . . . (p. 46)

A topical organization of the literature review means that reference may be made in different places to the same study; but such an arrangement gives the reader a systematic, complete picture of the "state of the art" and shows more clearly how the present investigation fits into the picture and how it helps provide the missing knowledge than does a chronological or an alphabetic organization. The review of the literature then becomes an integral part of the study and not merely padding in the study.

One should guard against padding the study with studies that are only remotely related to the topic under consideration. Only studies that are relevant and that set the stage for one's study should be included. For instance, if one's study involves the effect of immediate versus delayed reinforcement in classes composed of "normal" high school students, one would likely exclude studies of these effects on mentally retarded elementary school children.

If the past research concerning a topic is very extensive and if several studies have covered essentially the same ground, these can quite conveniently be organized in a table as in Illustration 5-1, excerpted from the study be Ford and Patterson (1967) of problems of student teachers. The major points of comparison and contrast among the studies can then be discussed in the text.

A Comparative Summary of Problems Identifed in Studies Relating to Student Teaching in Business Education

Problems Identified in Present Study	Previous Investigators				
	Bell-1952	Gress-1952	Burras-1956	Polson-1960	Theune-1960
Discipline	x	x	x	x	x
Classroom organization and management		x	x		x
Methods and techniques	x	x			
Individual differences				x	x
Motivation				x	
Knowledge of subject matter					
Selection and organization of subject matter	x				
Presentation of subject matter to students	x	x	x	x	x
Testing and evaluation	x	x	x	x	x
Assignments					
Personal Problems					

Illustration 5-1. Using a Table to Organize the Review of the Literature.

One study should not be summarized and then another and another rather like an extensive, detailed annotated bibliography; nor should the review of the literature be a series of quotations connected by a few words by the researcher. Quotations should be used only when the individual whose work is being reviewed has uniquely expressed an idea that no other wording would convey. Paraphrasing is much preferred to paragraph after paragraph of quotations. Some graduate schools require that dissertation writers get written permission from the author to quote passages of more than 25 or 50 words.

By using a topical organization and pointing out strengths and weaknesses of the studies reviewed that form the basis of decisions regarding one's own study, the review of the literature serves a real purpose.

Describing the Procedures

The procedures section is perhaps the crux of the research report; it is the background against which the reader evaluates the findings and the conclusions. The account of the procedures allows the reader to decide whether the findings are logical and whether the procedures really could lead to the findings and conclusions stated by the investigator. Therefore, **all** steps in the procedures must be included in detail so that the reader will know exactly how the procedures were carried out. The details should be complete enough so that another researcher following the described procedures could obtain similar results.

Selecting the Sample

The reader should have no question about how the sample was selected from the specified population—whether, for instance, the population was all the superintendents of secondary schools of a certain size in a specified geographic area, or whether it was the students in six classes of eleventh grade bookkeeping. The procedures section is the appropriate place for the description of the sample and the details about the sample or population selection. The population can be any group of subjects that the investigator designates: it may be a small population such as the 50 state supervisors of business and office education, or it may be large enough to encompass all secondary schools in the United States.

If a sample is selected from a population to provide the data for the study, great care must be taken in the research report in detailing the exact sampling procedures used. Simply saying that a random sample was selected is not sufficient; the method by which randomization was carried out must be explained, whether by use of a table of random numbers (and which one), by drawing numbered slips of paper from a hat, or by selecting from a list constituting the population every **nth** name starting with the **xth** one, **x** having been randomly selected from the first **n** names. If a convenience sample is used (the first 100 people one meets who have red hair, for example, or the closest ninth grade students), the method of and reason for their selection must be clearly explained. The size of both the population and the sample must be given, along with the basis of the decision regarding the sample size.

The same kind of detailed information must be supplied in the report regarding the returns from a survey study. The percentage of returns must be given, along with the number of questionnaires sent out or interviews or observations planned, as well as the number of unusable responses. These data are part of the description of the sample in the procedures chapter and **should not be considered findings.**

These descriptive data about the respondents usually are the basis of classification of the data pertaining directly to the questions under study and are not findings of the study. Data concerning the respondents' ages, income brackets, or educational levels may be collected on the survey instrument on which the main questions are asked. These demographic data do not pertain to the main question, so they should be reported as a part of the description of the sample in the procedures section rather than as a part of the findings.

The findings chapter should include only the replies to questions that directly relate to the questions that are part of the problem. For instance, in a study of opinions of bookkeeping teachers regarding certain instructional procedures, one of the questions might be whether or not the teachers feel that all calculations in bookkeeping classes should be done on machines rather than by hand. The responses might be reported first for the total group of respondents and then by groups of teachers classified according to the years of experience or according to whether male or female teachers made the responses. Although the data regarding teaching experience and the teachers' sex were collected on the questionnaire, the problem itself was not concerned with the characteristics of these teachers but with their opinions regarding the use of calculating machines in bookkeeping classes. Thus descriptions of the distribution of the respondents according to experience and sex is properly reported in that portion of the procedures section dealing with the sample. The fact that 123 of the respondents were men and 156 were women is not a **finding** regarding the question about the use of machines in bookkeeping; it merely describes the sample.

Procedures in Experimental Studies

Experimental and prediction procedures must be described in great detail. Every facet of the experimental procedures must be spelled out precisely and explicitly. Of particular importance are the selection and size of the sample, the assignment of the treatments to the subjects (or vice versa), the exact experimental procedures employed, the controls used, and the selection or development of all measures used.

As previously pointed out, the sampling techniques are a crucial aspect of a research study and must be spelled out carefully so that the readers will know how broadly and accurately they may generalize from the sample selected. But even more important is the random assignment of the subjects to the experimental and control groups and the random assignment of the different experimental procedures to the groups. Such detail is necessary for the reader to be aware that the investigator recognized the necessity of controlling as many variables as possible so that outcomes may be attributed to the experimental procedures rather than to any of a number of other factors.

Any other control measures must be described fully also. Were certain students eliminated from the study? Why? Were the teaching materials used the same in all groups, or were they the experimental variable? Were the ages of the students comparable? Were the abilities of the students in the groups the same, and if not, what procedures were used to equate the groups? Or were the students grouped so that interactions between groups and treatments could be studied? What factors other than the experimental variables may have been expected to influence the possible outcomes? All such questions must be answered explicitly.

A detailed explanation, including justification, of each step of the experimental procedures is essential in a well-written report. This detail allows the reader to decide whether the procedures as carried out could lead to the answers to the questions proposed in the study. These procedures, of course, would include such factors as any pretesting of the groups, the purpose of the pretesting, and the reasons the particular measures were used.

Although no experimental or prediction study is any better than the criterion measures used, the complete description of all measuring instruments is frequently neglected in research reports. In many studies only a brief statement is made to the effect that at the end of the experiment the students were given a test prepared by the investigator. The least that should be done is to include in the appendix any test or other instrument if it is not a published test available from a commercial source. If the instrument was developed specifically for the study, all details about how it was constructed should be written into the report. Such details should include information regarding any tryouts of the test, the subjects and how comparable they were to the subjects of the study itself, an indication that some preliminary testing of the items was done and that some item analysis was performed, as well as specific techniques used for determining validity and reliability. Whether reliability was established by one of the Kuder-Richardson formulas (specify which one), a split-half technique, or a test-retest procedure, the resulting reliability coefficient must be reported.

If published tests are used as the criterion measures, the technical data regarding validity and reliability must be reported. Without such information about the criterion measures, the reader has no idea whether or not the findings and conclusions of the study can be relied upon, for if the measurements are not dependable or if they are not suitable for the purpose of the study, the whole project is of doubtful value.

The detailed classroom procedures for each of the participating groups, the teacher's part in the experiment, and the time spent on various activities are but a few of the other essential details that must be included in a good description of the procedures.

Procedures in Descriptive Studies

Sufficient detail of the procedures used is also important in descriptive studies, but it may differ slightly from that in experimental studies. The procedures section in a descriptive study may be less complex than in an experimental study, but the detail is still as important. The selection of the population and sample and the development of the data-gathering instruments or techniques are especially important.

Defining the appropriate population is of utmost importance, but the population should not be confused with the sample, as in one study where the statement was made that "the population of this study was a sample of...." One surveys either a population or a sample, and the researcher must make clear in his report which he is doing. If the selection procedures are carefully delineated, the reader can then generalize the conclusions properly.

Along with the description of the sampling procedures, a complete description of the respondents is required. The researcher must specify first the exact number of questionnaires mailed out, for example, or the number of interviews planned. He should then report the actual number of responses or interviews and how many of the responses were complete and usable. The number of usable returns should be reported in relation to the total number of questionnaires sent and not as a percentage of the number of returns. In many studies the percentage of return is never stated precisely, and the reader must ferret out this information from the data presented in the findings. Knowing the number of responses in relation to the number of data-gathering instruments sent, preferably stated as a percentage, is essential for the reader to have an estimate of the reliability of the findings and of their general applicability.

A very important part of the description of the sample and the respondents is a comparison of the characteristics of the respondents with those of the complete sample or population (West, 1963). Certain characteristics of the persons making up the original sample may be known or available to the investigator (e.g., the number of clerical employees in each of the SIC categories). If a sample of clerical employees is being surveyed and the sample was selected in proportion to the number of such employees in each SIC category, the writer of the research report should report the number of returns obtained from persons in each of the categories and compare the distribution of the returns in the categories with the distribution of the population (preferably with some statistical test as X^2). This is necessary so that the reader may have some estimate of the **representativeness** of the respondents, [illegible] in establishing more clearly the credibility of the findings. Of course, the variable on which the responses may be classified must be ones that form subgroups that are likely to have different responses. If a particular classification is not likely to produce differences in responses, then avoid wasting time by making such classifications.

Nonrespondents tend to be unlike those who respond to inquiries on questionnaires; therefore, every effort should be made to get some information from a sample of the nonrespondents by which they may be compared with the respondents. If this desirable procedure has been followed, it should be carefully spelled out in the report. Below is an example of such a technique. Although the procedures could have been explained a little more fully, the brief description given does indicate to the reader something of the nature of the nonrespondents.

> As indicated, 56 percent of the business firms returned either usable or unusable questionnaires, indicating that 44 percent of the firms did not respond to any of the three mailings. In order to determine that the 44 percent did not represent a source of data different from the 56 percent, the following procedures were followed. A total of ten of the nonresponding firms were selected at random and contacted via telephone. The results of these conversations indicated that the non-responding firms were not substantively different from the responding firms.

Readers would have a great deal more information if they knew in what ways the firms did not differ: number of employees? sales volume? percent of employees who are union members? or on their responses to two or three key questions? The bases of comparison of the respondents and the nonrespondents should be reported so the reader can evaluate their relevance to the study.

As in reporting all other procedures, the basis of each decision should be reported, for instance, about how the data were gathered and whether a published, standardized instrument was used or whether some type of questionnaire or rating scale was developed by the investigator. In a report involving the use of some such instrument as a semantic differential or a Likert- or a Guttman-type scale, techniques that may not be familiar to many readers, additional detail regarding the technique, its use, and the underlying theory should be included in the report. The reader will thus be able to understand the meaning of the numbers that represent the direction and degree of agreement of the opinions being measured, enabling him to interpret properly the findings of the study.

Pilot Studies

In most experimental research and in many other kinds of studies, too, certain procedures may be novel to the participants, so researchers need to try out the procedure to see if it is feasible and manageable in classrooms or to locate troublesome points in following the directions on questionnaires or other data-gathering instruments. Therefore, usually a pilot study is necessary and well worth the time and effort. Any pilot study and its problems must be reported at least briefly, pointing out any difficulties encountered and how they led to changes in the procedure. This adds strength to the study by showing the reader that the investigator was aware of the possibility of outside variables influencing the outcomes and that the procedures used in the actual study itself were sound and were handled without difficulty.

In other words, every step and every decision made in planning a study must be accounted for in the research report. Each decision should be supported by either a practical or theoretical base; the investigator should show exactly why the decision was made to use a given procedure rather than an alternative, especially if the alternative is one that is known to many people. For example, in several shorthand experiments done in first-semester shorthand classes, the criterion score has been only the number of words transcribed correctly on a dictation test. This seems to be a logical criterion measure for beginning shorthand. In some studies, however, missing or misused punctuation, misspellings, or faulty paragraphing are counted as errors. This is done even though no reason is given for such a practice and in spite of the fact that the dictation tests were designed to measure basic skill in recording dictation and not to measure English skills. Counting punctuation errors is the prerogative of the researcher; but in order to evaluate the outcomes of the study, the reader needs to know why English grammar and mechanics errors were counted when mailability was not part of the study.

In a bookkeeping experiment, a researcher may have decided to equate the groups using as one of the equating factors a measure of interest. On what basis was such a decision made? What evidence does the researcher have that interest of the student affects learning in bookkeeping to such an extent that it should be a factor to control? The investigator must justify every decision made regarding any aspect of the study, showing the logic of the decisions based on his own observations or on past research which suggests that such a decision ought to be made. Whatever the basis for each decision, it should be explained in the report. A pilot study may have helped the researcher decide which instrument to use to measure interest.

Thus in writing the report the investigator, without insulting the intelligence of the reader, must present in great detail every step in the procedure, discussing the basis on which each decision was made. Such decisions will involve everything one has done by way of controlling the experimental conditions, deciding on the actual procedures to be used by the subjects of the experiment, and determining the means of measuring the outcomes. In addition, one must be as explicit as possible, of course, in describing the selection of the sample.

The procedures section of an experimental or correlational study is often quite lengthy, since much preliminary planning and preparation are usually necessary in conducting these studies. However, Dvorak (1971, p. 341) cautions the writer against including the simple details known by anyone conversant with such research methods. In the selection of a random sample using a table of random numbers, for example, one need not go into detail about whether the researcher proceeded horizontally or vertically from a random spot on the page.

Reporting the Findings

The findings should, of course, be presented in a lucid, interesting way. Frequently, because of the similarity of a series of related findings, such as those from a rather lengthy questionnaire, writing the findings section in a way that is not monotonous is indeed a challenging task. One must try to vary the wording for each section of findings or perhaps use different means of presenting data that may be appropriate—tables, graphs, figures, or diagrams, as well as the narrative.

Tables

Most findings lend themselves to presentation in a table. The purpose of organizing the findings into tables is to reduce the amount of writing necessary to convey the information; thus, stating in the text every number presented in a table is unnecessary. The narrative accompanying each table should point up any trends, contrasts, or seemingly inconsistent figures in the table. In one study a table summarizing the formal educational background of students enrolled in business schools contained a figure of -1 under a column heading "% Did Not Complete High School." No explanation was given in the report of how the -1 was obtained or of its meaning. Any such questionable figure must be fully explained.

The research should also make comparisons among the groups about whom the data are being presented or point out the contrasts from one classification of data to another. The researcher must be aware of all the information in the table in order to make comparisons and contrasts. Sometimes looking at two or three lines of the data together makes an important contrast rather than just looking at each line separately. What do you see in the following table, taken from Lee's (1970) study comparing Certified Professional Secretaries in Louisiana with non-Certified Professional Secretaries who were members of the National Secretaries Association?

Table 5-1. Aspect of their Job Liked Best by Louisiana CPSs and Non-CPSs

Aspect liked best	CPSs		Non-CPSs	
	Number	Percent	Number	Percent
Responsibility	17	32.7	29	18.8
Customer and client relationship	8	15.4	2	1.3
Working conditions—physical plant	7	13.5	21	13.6
Challenge	7	13.5	15	9.7
Variety	6	11.5	38	24.7
Meet many people	5	9.6	7	4.6
Office work	1	1.9	7	4.6
Other	1	1.9	16	10.4
No response	0	0.0	19	12.3
Total	52	100.0	154	100.0

Source: Lee, L. W. *Personality Characteristics and Job Satisfactions of Certified Professional Secretaries and Non-Certified Secretaries.* Doctoral dissertation, Greeley, CO: University of Northern Colorado, 1970, p. 117.

Of course, one sees immediately that what CPSs seem to like most about their jobs is the responsibility entailed in it, while the non-CPSs mention variety as being the aspect of the job they like best. Both groups mentioned physical working conditions about as frequently as the other.

But another way we can look at those same figures is by considering various percentages in relation to each other. We might say that CPSs reported nearly twice as often as non-CPSs that they liked the responsibility connected with their jobs, but variety in their jobs was reported about twice as frequently (25 percent) by non-CPSs as by CPSs (12 percent) as the most favorable aspect of their jobs. On the other hand, about an equal proportion of both groups said that physical working conditions were the things they liked best.

Also, if certain categories of responses are combined, the contrast may be more pronounced or meaningful; for instance, contrast the percentage of the two groups when both lines concerned with working with people are combined: "Customer and Client Relations" and "Meet Many People" account for 25 percent of the CPSs' responses, but only about 6 percent of the non-CPSs' responses. Notice in the quoted paragraph below that Lee not only discusses the differences between the two groups, but she also explains why this finding regarding meeting people is a logical finding:

> Some similarities were found between the two groups of secretaries concerning facets of the job they liked best. The five things they liked best about their jobs, in order of frequency of mention, were responsibility, the customer and client relationship, working conditions, challenge, and variety. Of these elements liked best by CPSs, non-CPSs listed four: variety, responsibility, working conditions, and challenge.
>
> The greatest difference between the groups existed in the job satisfier of customer and client relationship, which seemed an important satisfaction to 15.4 percent of CPSs and to only about 1 percent of non-CPSs. When customer and client relationship was combined with meeting people, an even greater difference existed; this combination of job satisfiers was mentioned by 25 percent of the CPSs as compared to about 5.9 percent of the non-CPSs. A logical explanation might be that a larger proportion of the CPSs (27 percent) as compared to non-CPSs (2 percent) as shown in Table 4 [Here Lee is referring to a table occurring earlier in the study.—MH] held titles of executive secretary or administrative assistant; and that these CPSs, who occupy a higher level in the management team hierarchy than do non-CPSs, are generally more concerned with public relations. (p. 118)

The same data sometimes need to be presented in different ways in separate tables to give the reader all the information contained in the data. Tables 5-2 and 5-3 taken from Ober's (1974) study of typing tasks performed by beginning typists illustrate this point. In which of the two tables will you find out to what extent different kinds of typing tasks are performed in different types of business? Which of the tables tells you this: Of the documents that were analyzed, which types of working papers came predominantly from which types of business? Which table would be most helpful to a student who wants to know the kinds of typing tasks most likely to be performed in a particular kind of business?

Table 5-2. Kinds of Items Typed in Business Offices, Classifed by Industry

Industry		Kind of item						
		Forms	Letters	Memos	Reports	Tables	Other	Totals
Manufacturing	No.	64	25	21	2			112
	%	28.8%	14.5%	27.2%	9.5%			21.1%
Trade	No.	14	30	14			2	60
	%	6.3%	17.3%	18.2%			11.1%	11.3%
Insurance/Banking	No.	27	30	11	4	2		74
	%	12.2%	17.3%	14.3%	19.0%	10.0%		13.9%
Education	No.	26	13	10	8	4	4	65
	%	11.7%	7.5%	13.0%	38.1%	20.0%	22.2%	12.2%
Services	No.	61	53	13	4	4	10	145
	%	27.5%	30.6%	16.9%	19.0%	20.0%	55.6%	27.3%
Miscellaneous	No.	30	22	8	3	10	2	75
	%	13.5%	12.7%	10.4%	14.3%	50.0%	11.1%	14.1%
Totals	No.	222	173	77	21	20	18	531
	%	41.8%	32.6%	14.5%	4.0%	3.8%	3.4%	99.9%

Source: Ober, B. Scot. *An Analysis of the Business Working Papers Typed by Beginning Office Workers.* Doctoral dissertation, Columbus, OH: The Ohio State University, 1974, p. 61.

Table 5-3. Makeup of Industry Typing, Classified by Kind of Item

Kind of Item		Industry						
		Manuf.	Trade	Ins./Bank	Educ.	Services	Misc.	Totals
Forms	No.	64	14	27	26	61	30	222
	%	57.1%	23.3%	36.5%	40.0%	42.1%	40.0%	41.8%
Letters	No.	25	30	30	13	53	22	173
	%	22.3%	50.0%	40.5%	20.0%	36.6%	29.3%	32.6%
Memorandums	No.	21	14	11	10	13	8	77
	%	18.8%	23.3%	14.9%	15.4%	9.0%	10.7%	14.5%
Reports	No.	2		4	8	4	3	21
	%	1.8%		5.4%	12.3%	2.8%	4.0%	4.0%
Tables	No.			2	4	4	10	20
	%			2.7%	6.2%	2.8%	13.3%	3.8%
Other	No.		2		4	10	2	18
	%		3.3%		6.2%	6.9%	2.7%	3.4%
Totals	No.	112	60	74	65	145	75	531
	%	21.1%	11.3%	13.9%	12.2%	27.3%	14.1%	100.0%

Source: Ober, B. Scot, *An Analysis of the Business Working Papers Typed by Beginning Office Workers,* Doctora
dissertation, Columbus, OH: The Ohio State University, 1974, p. 63.

In Table 5-2 Ober shows the extent to which the documents he analyzed came from various types of business. For instance, of all the forms analyzed, 56 percent came from only two categories of business, manufacturing and services. More of the letters came from service industries than from any other business group. Looking at these same data from another angle, though, may be more helpful to teachers and students wanting to know the kinds of typing tasks done by beginning office workers in the various types of business. Ober presented that breakdown of the data in Table 5-3. There we see that more of the typing tasks in manufacturing firms involved typing forms than anything else; in fact, forms typing was done more than all other tasks combined, accounting for over half the work of beginning office workers. On the other hand, typists in the trades segment of business typed more letters than any other kind of document, with the rest of their work being typing forms and memos, in about equal proportions. Forms and letters comprised over three-fourths of the work of the beginning office workers in banks and insurance offices.

Thus, in examining and analyzing data in tables, the researcher must look at them from all possible viewpoints to get the most out of the data. One has to "mine" the data by examining the figures thoroughly and organizing and analyzing them in different ways. Much valuable data may be lost because the researcher has examined only one dimension of the tables and has failed to consider two or more categories of data together.

Tables should be presented and titled in such a manner that one need not read the text to understand them; they should be self-explanatory. The title should make completely clear what the table is about so that the reader need not refer to the text to know what the figures represent. The units should be clearly designated, whether they represent money, pupils, words per minute or whatever. The reader should not have to read the explanation of the table in order to grasp the meaning of the numbers.

Graphs

Some data may be more meaningfully presented by means of graphs. Comparisons, trends, or contrasts are more clearly seen in graphs than they are in tables of numbers. Trends are more clearly shown by line graphs while contrasts may show up more readily in bar graphs. To show relationships of parts to a whole, a pie chart may be the choice, provided it doesn't have too many divisions. Still other types of charts may be more appropriate for presenting other kinds of data; namely, surface charts, sliding bar charts, or paired or grouped bars or column charts, among others (Spear, 1975).

Any graph, however, must be appropriate for displaying the particular type of data one is presenting in order to convey an undistorted message. If you wish to show a trend in total enrollments in business education classes for the past ten years, you would choose a line graph or perhaps a vertical bar (column) graph. The vertical bars will emphasize the years more than the line graph does, while the line will emphasize the trend over the years. On the other hand, if you wish to show enrollments in different types of business classes in one particular year, a bar graph (usually horizontal bars) would be appropriate. However, if you want to compare the enrollments over a period of years of boys and girls in one of the business classes, you could use grouped bars. In such a case the group consists of only two bars, one for boys and one for girls. It would be similar to the ones showing school bus expenditures in Figures 5-1A and 5-1B. On the other hand, if you have two sets of data for each year, each requiring a different scale to compare the items, you could use paired bars as in Figure 5-2.

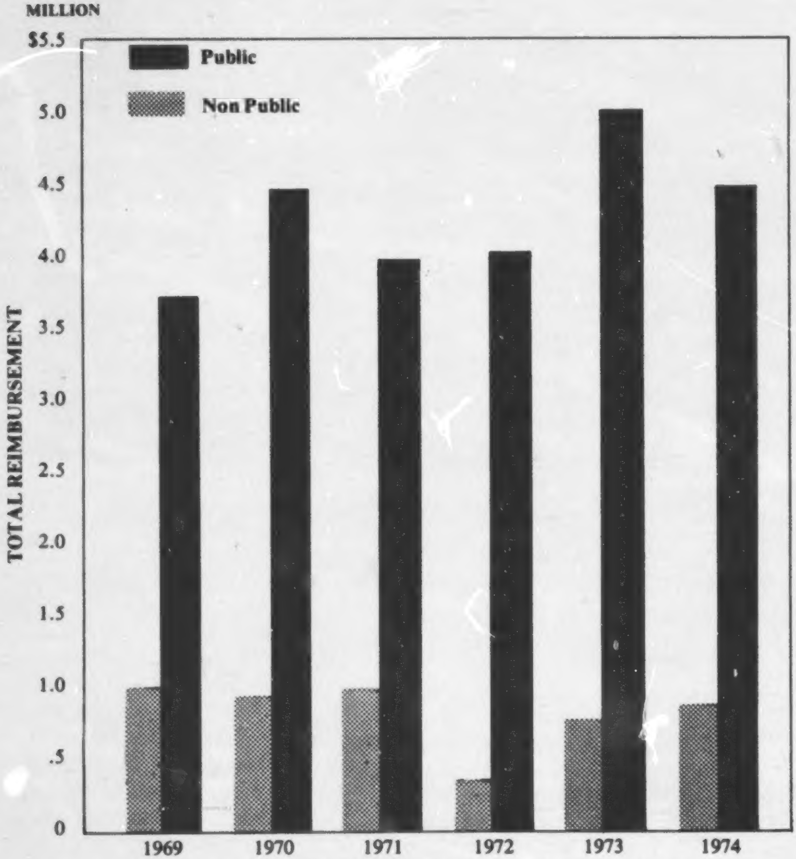


Figure 5-1A. Total Reimbursement for School Buses.

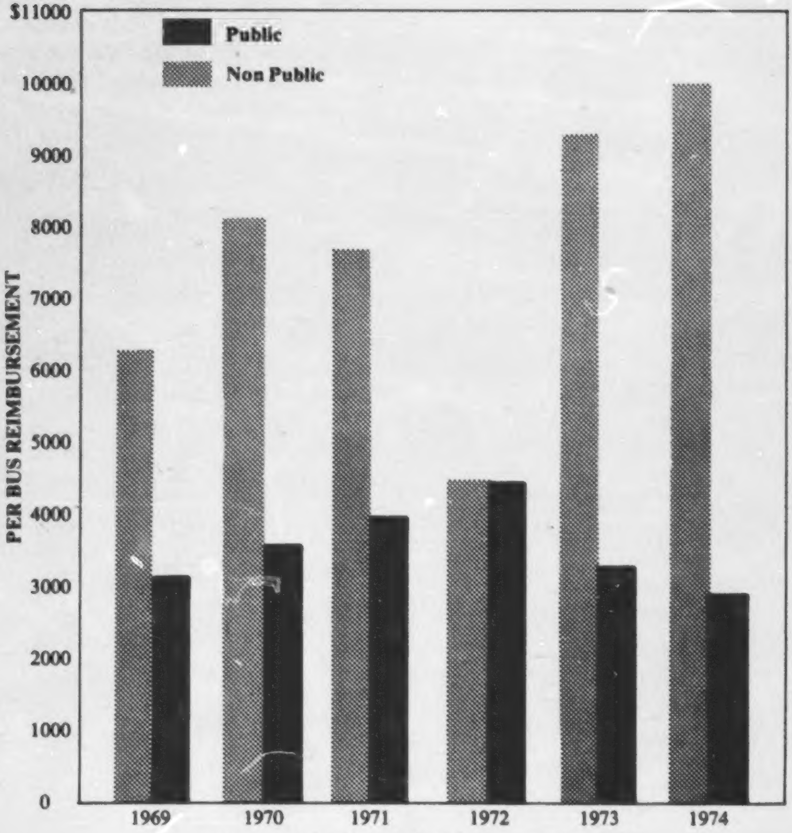


Figure 5-1B. State Reimbursement Per Bus.

Following certain "rules" for constructing graphs will insure that the resulting chart will "tell the truth" and not distort the data; that it will give the appropriate message.

1. **Use the proper scale.** Expanding or contracting the scale in one direction without corresponding adjustments on the other can change the "message" the graph conveys (Figure 5-3). What is your impression of Graphs B and C in comparison with A? These three graphs present the same data, but the scales have been changed. Figure A is the original data; in B the horizontal scale has been reduced by half, with the vertical scale unaltered. On the other hand, in Graph C the vertical scale is only half

as large as in A, with the horizontal scale the same as in Figure A. The rate of growth seems to be more rapid in Graph B and slower in Graph C than it was in the original presentation. Changing the scale changes the "message."

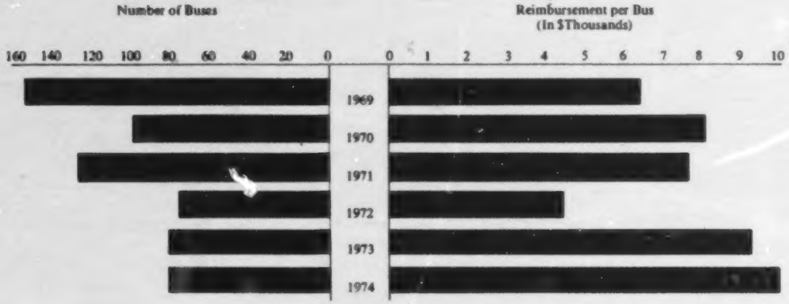


Figure 5-2. State Reimbursement for School Bus Purchase.

2. **Include the base line.** Also, in the original graph from which the graphs in Figure 5-3 were taken (A), the base line is labeled 50% and the top line is 80%. This is another way of distorting the scale and the message and one that is often used to exaggerate a point.

Compare Figure 5-3A with Figure 5-4. The size of both graphs is the same, but Figure 5-4 shows the entire range of percentages from 0. Again, using the top end of the scale conveys a different message than using the complete scale.

3. **Use the same sized interval for each unit of data.** In Figure 5-5A the first date on the horizontal axis is 1964 and the next one is 1969, a span of five years. The remaining intervals, which are the same size as the first interval, represent only one year apiece. In Figure 5-5A, the message is that enrollments in vocational classes increased at a rapid rate right after enactment of the 1963 Vocational Education Act.

But when space is allowed for each of the five years between 1964 and 1969, the trend line does not rise nearly as sharply, giving quite a different picture of the enrollment trend (Figure 5-5B).

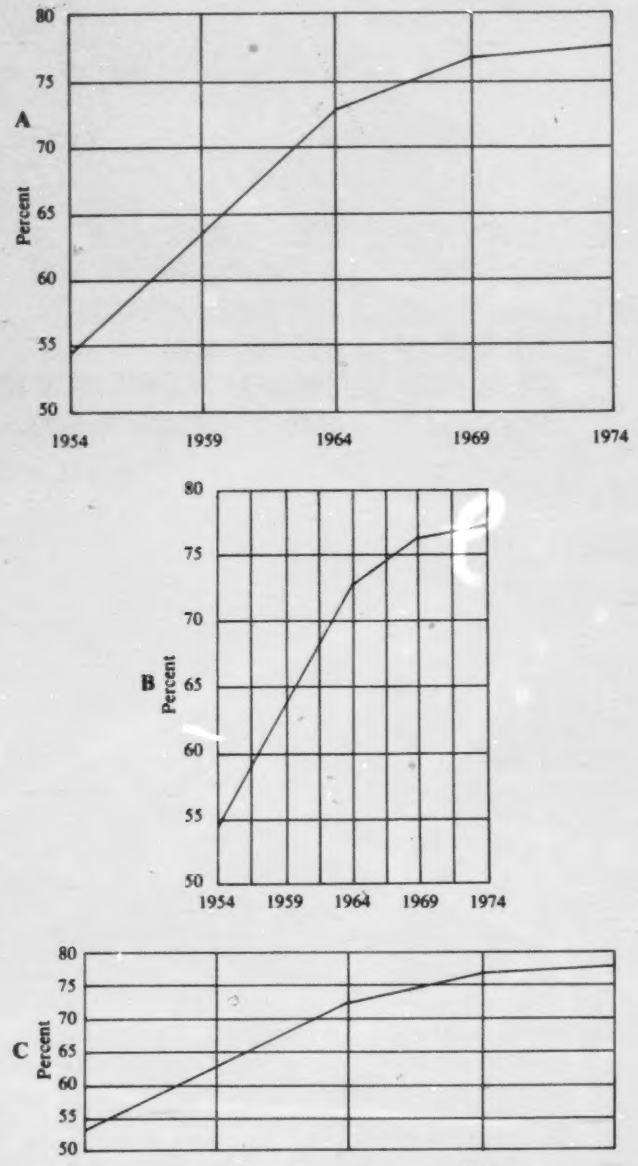


Figure 5-3. Effect of Changing the Grid Scales.

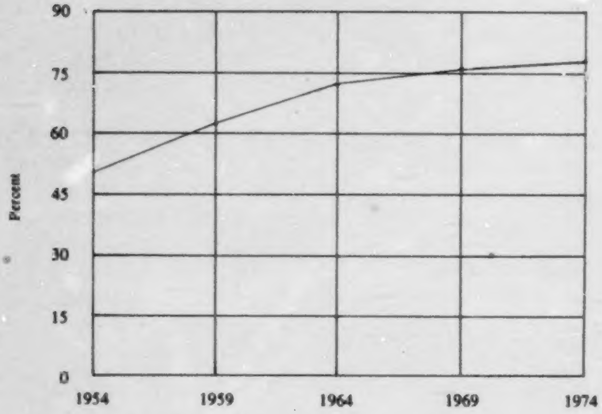
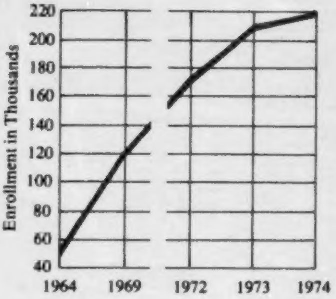


Figure 5-4. Effect of Including the Entire Range of Data.

A

Year	Total
1964	54,826
1969	122,518
1972	175,018
1973	211,845
1974	218,590



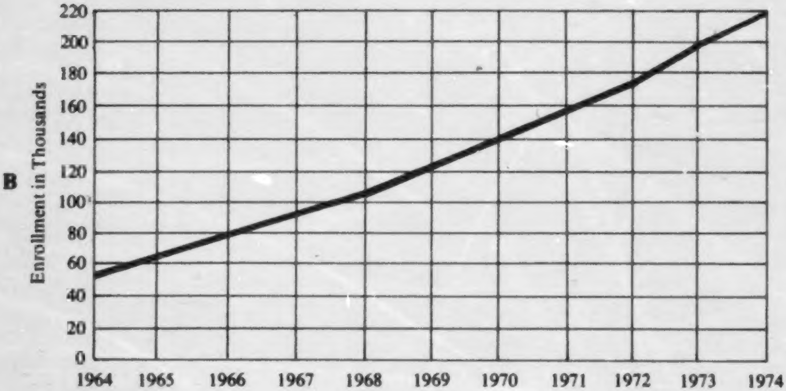


Figure 5-5. High School Vocational Enrollment, 1964-1974.

4. **Do not break a graph.** Not only does Figure 5-5A have incorrect unit representations, it breaks the presentation between 1969 and 1971, increasing the distortion even more, and it gives a different impression than the data indicate. Figure 5-5B shows that when all the years are accounted for, the general trend line indicates a much slower rate of increase than appeared in Figure 5-5A. Breaking a bar or two on a bar chart may be appropriate if the break is made beyond the next longest bar and if only one or two of the bars are so much longer than the rest that the shortest bar would nearly "disappear" if the entire longest bar were shown to scale. (Spear, 1969, pp. 58-60)

Year	Number	Percent of Total Graduates
1964	3,443	3.1
1969	3,795	2.7
1972	3,855	2.6
1973	5,017	3.3
1974	4,364	2.8

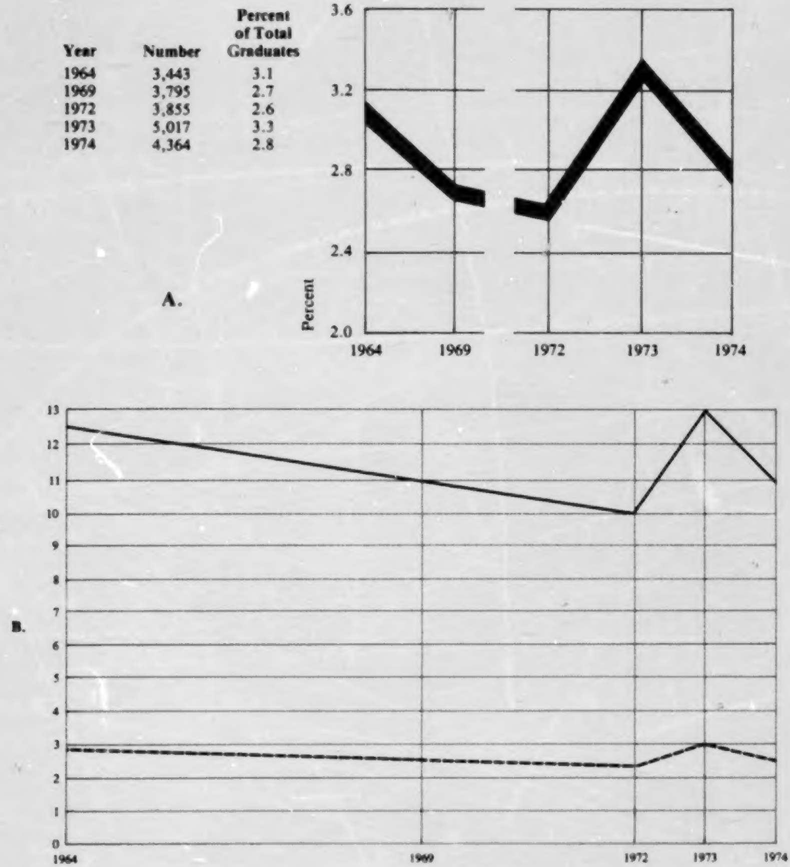


Figure 5-6. Graduates in Top 25 Percent of Class Unable to Attend College.

5. **Properly label the chart.** The caption in Figure 5-6 implies that the percentages reported are of the top 25 percent of the students who do not attend college. However, looking carefully at the data in the table accompanying the graph, one sees that the percentages are of the **total** number of graduates and not of the top 25 percent of the graduating class; thus, the small percentages of 2.5 to 3 percent create (as shown in Figure 5-6A) the impression that almost all the students in the top 25 percent of their graduating classes do go on to college.

If the percentage is taken of the number of students in the top 25 percent of the class, as the table title implies, the graph will, of course, have the same trend, but the figures are more "truthful." Calculating the percentages that way, they range from 10 to 13 percent—quite a different story—shown by the solid line in Figure 5-6B. The broken line in Figure 5-6B is for the data based on the total number of graduates. Notice also the unequal intervals in Figure 5-6A, and how different the trend appears in Figure 5-6B.

6. **Present what is in the data.** What do you see in the data in Table 5-4?

Table 5-4. State Reimbursement for School Bus Purchase.

	Buses		Reimbursement
1969	154	Nonpublic	$ 981,933
	1,162	Public	3,688,439
1970	115	Nonpublic	932,508
	1,148	Public	4,449,444
1971	124	Nonpublic	947,268
	992	Public	3,987,001
1972	75	Nonpublic	333,628
	900	Public	4,003,524
1973	81	Nonpublic	752,915
	1,329	Public	4,984,324
1974	81	Nonpublic	803,966
	1,210	Public	4,400,054

Source: *Annual Report,* 1974, State Board of Education, Columbus, OH, p. 36.

Obviously the data may be presented in a graph as it was in Figure 5-1A, showing a great deal more reimbursement being made to public than to nonpublic schools for buses. But look more closely at the data, taking into consideration the number of school buses for which reimbursement was made. If the mean reimbursement per bus is shown for each type of school system, the picture of the school bus expenditure looks quite different from the original graph (Figure 5-1A and B).

Such data presentation points up the need for alertness to what is actually in tables and that the way the data are displayed may create an illusion. Close attention to the detail in a table, looking for relationships among the numbers and examining them in different ways is essential in reading and in displaying data.

Discussing the Findings

Any seemingly inconsistent findings must be discussed thoroughly in light of any limitations in the data or in the procedures that may account for the inconsistencies. For instance, if in an experiment conducted in several classes, results from all the classes but one were in favor of the experimental procedure and in this one class the outcomes were exactly opposite to an extreme degree, the researcher should carefully examine the data and procedures for any possible reasons for the inconsistent findings. If the students had been randomly assigned to the treatment groups, could all the most able students have been assigned by chance to the control group? Did a different teacher teach the experimental and control groups? Did a student teacher or substitute teacher who was unfamiliar with the experimental procedure conduct classes some of the time? Was the timing of tests incorrect in one of the classes? Any of a number of other factors may have accounted for the result, and they must be brought to the attention of the reader.

Among the reported findings in a study (Boak & Conklin, 1975) of the effect of teacher interpersonal relations and junior high school students' achievement and anxiety is this statement:

> Approximately 8½ percent of the variance in reading comprehension scores was accounted for by teacher levels of interpersonal skills. No appreciable variance in vocabulary or anxiety scores was accounted for by these skills (the effects of the other variables having been removed). (p. 540)

The researchers then speculate why only this small amount of the variance in achievement could be accounted for by teacher interpersonal relationships at the junior high school level:

> The results reported lend some support to the importance of teacher interpersonal skills in the classroom, in that higher levels were significantly related to greater gains in achievement for the language arts (grade eight) and mathematics groups. However, while not disputing the significance of the presented results or similar results of other researchers, definite questions arise upon looking at them from a measure of association between student achievement and the two levels of teacher interpersonal skills.
>
> Why did interpersonal skills account for only 2 to 8½ percent of the variance in student achievement? Although the "high" and "low" groups of teachers were significantly different ($p = .001$), the mean level of functioning of the high group was below minimally facilitative levels as defined by Carkhuff (1969) and thus could be expected to have only limited effects on student achievement. It is perhaps more important to consider that the "high" groups accounted for as much of the variance in student achievement as they did, especially since the teachers in this study spent only one period per day with each of their classes. Therefore, students interacted with several teachers each day and the facilitative effects of one teacher may have been nullified by the others or by parents. (p. 541)

Boak and Conklin then discuss another of their findings in this same way, bringing up a probable limitation of one of their measuring instruments:

> There was a high degree of consistency between pre- and post-anxiety scores for all groups of students. Either the instrument was not sensitive enough to detect a change in student anxiety over a six-month period, or student anxiety was relatively stable. The mean anxiety scores indicated that, overall, the students tended to be well adjusted. (p. 542)

One must also read the findings or present them in light of the samples of subjects supplying the data. Assume that the findings of a survey of data processing employees, made to discover which courses in the high school business curriculum they considered the most valuable, seemed to indicate that typewriting was the most essential course for working in the data processing department of a firm. An investigation of the replies might reveal that a disproportionate number of responses came from keypunch operators, who use typing skill constantly in the performance of their jobs. In writing and in reading the research report one must be very careful to take into account any limitations of the data in terms of the sample, of the instrument, of the proportion of responses from various subgroups within the sample, and of the characteristics of the respondents.

Some researchers feel that if their experiment has not produced a significant difference, they have not made a contribution to the field and that they really have nothing to conclude except that no differences existed between the groups. This feeling is reflected in several studies. Such findings, though, should be discussed as thoroughly and their meaning explored as deeply as the findings of a study yielding significant differences. Nonsignificant differences are a contribution since they show that the experimental method under study is another way to approach teaching the subject matter under consideration and that teachers who are looking for a different way of teaching may use it knowing that it will not be detrimental to the students' progress.

Also, the findings in prediction studies showing that the variables being studied do not contribute to the prediction of the criterion measure should be thoroughly discussed. Findings such as these are important for any person in the future who may wish to conduct a prediction study in the same field. The next researcher to tackle the problem will know that he need not be concerned with those particular measures and might more profitably explore other likely predictors (provided, of course, that the criterion measures were valid measures of "success"). Nonsignificant findings are, indeed, often as important as those yielding significant differences and should be treated thoroughly in writing the report.

Summarizing, Concluding, and Recommending

The last chapter of the research report will usually include a summary of the study along with the conclusions and any recommendations the reseacher may wish to make.

Summarizing

The summary serves as a synopsis, which many individuals will read first in order to see if the study is worth further, more thorough reading. The summary, although complete, should be just that—a summary and not a repetition of sections of the previous chapters. The procedures should be summarized in general terms, with only enough detail given for the reader to obtain a general picture of what was done, yet with sufficient detail so that he will know if he wants to read the entire study. The summary of the findings should also be in synopsis form and should not be a duplicate of the findings as given in that chapter. Summary tables are a convenient, meaningful, efficient way to present findings in this part of the report, especially in summarizing results of a questionnaire or other survey or descriptive study. The summary of the findings, however, must be complete and presented in such a way that the findings lead the reader naturally to the conclusions.

Drawing Conclusions

The **conclusions** one draws must **derive directly from the reported findings.** Because of this, the investigator must have carefully presented in detail the procedures so that the reader can see that the findings are logical and that, in turn, the conclusions are logically derived from the findings. The number of conclusions depends upon the questions asked in the statement of the problem. Most studies, of necessity, are limited to one or two main problems. Although a whole series of subproblems may have been formulated in detailing the problem, this does not mean that a whole series of conclusions, one for each of the subproblems, is necessary, but rather that the findings related to each of the questions taken all together add up to one or two major conclusions regarding the main problem.

Hartman's (1970) study provides an example of this. He investigated the feasibility of combining beginning typing and beginning shorthand into a single course to be taught in one class period daily to students who had not studied either course before. The teacher in this experiment taught one combined class, one beginning shorthand class, and one beginning typing class—each for one period every day. Hartman posed six hypotheses regarding speed and accuracy scores in typing (one each on speed and accuracy on three typing tests) and six additional hypotheses regarding shorthand dictation skill (each based on a separate shorthand dictation test). Although he reported findings related to each of these six hypotheses, he drew only four conclusions regarding the effect of the experimental procedure on speed and accuracy in typewriting and on dictation skill in shorthand, and then he drew this one conclusion: Combining beginning shorthand and beginning typing into a single class was a feasible procedure.

Conclusions are not the same as findings and should not, therefore, be a restatement of the findings. In many studies comparing the relative effectiveness of two or more methods of teaching a particular course, an F value is reported significant at the .05 level. This is a perfectly legitimate finding and usually is well reported; but the researcher sometimes draws the conclusion relative to this particular finding that "a significant difference was found between the means in favor of..." The conclusion should not be that the means were significantly different, but rather something to the effect that Procedure A is likely to produce better results in the classroom than Procedure B, or that the use of Procedure A is feasible and is likely to help the students make more rapid progress than Procedure B.

This is not to say that findings are not mentioned in the conclusions section. Conclusions should be supported by the findings, but a separate conclusion need not be drawn for each finding. It is the **sum** of the evidence that leads to the conclusions, as illustrated by Osborn's (1970) study, in which she posed five questions in stating her problem. From the findings concerning each of the five questions, she drew one main conclusion and then from an incidental finding she drew a second conclusion. (The incidental finding was something that appeared in the data that had not been anticipated in planning the study.) Notice that Osborn first states the findings in very general terms and leads to the conclusions:

> Since the correlations between the scores on the shorthand symbol reading tests and the scores on the dictation tests were relatively high, the ability to read shorthand symbols in context does seem to have some influence on the ability of students to record and transcribe dictation. And, since the r's increase each time a "sense" is added (seeing only; seeing and hearing; and a combination of seeing, hearing, and feeling in the course of practicing their reading), the more senses involved in the

> learning process, the more important reading becomes in the ability to take and transcribe shorthand dictation. In addition, since the mean scores were highest (but not statistically significant) for the reading-listening-scribble-writing group and next highest for the reading-listening group on four of the six terminal dictation tests, the conclusion that the addition of an extra "sense" increases learning is further substantiated.
>
> Individual students tended to transcribe a similar number of words on all dictation tests (with the exception of Dictation Test 5, which is perhaps an invalid measure) regardless of the speed of dictation. The scores on the two 80 words a minute tests combined were slightly higher than those at 60 words a minute. Therefore, all speed tests in beginning shorthand classes might justifiably be dictated at the 80 words a minute rate. (pp. 83-84)

This last conclusion was supported by the findings from her own study, but Osborn should have also pointed out that her findings corroborated those in two other earlier studies in which the same findings were reported. A good research report will tie the findings of the present study to other studies that have explored the same problem. Wright (1975) does this in a discussion of the findings in a study of the affective and cognitive consequences of an open education elementary school. He points out how this study differs from some studies and corroborates others and discusses the importance of the findings in light of criticisms of the open classroom. He also explains, in part, the outcomes of his study:

> To assess the overall influence of this open education program, it is necessary to take careful note of these student variables which did not show significant differences for the two school groups. Contrary to the prevailing literature of this field (Haddon & Lytton, 1968, 1971; Roger, 1969; Weber, 1968), students in the open school did not have higher levels of tested creativity. Also, no open school had higher levels of tested creativity...
>
> The area of greatest impact associated with the implementation of this open education program was found with the various facets of pupil achievement. Specifically, after four and a half semesters in this environment, the students of the open school had significantly lower overall levels of academic achievement. These differences were especially noticeable in arithmetic computation, as well as in reading, and other reading related achievement areas.
>
> Rathbone (1972) has recently expressed the belief that the ultimate academic goals of open education should be the same as they are for the more traditional form of education. In many respects these goals are very conventional and involve the mastery of basic skills in such areas as reading and arithmetic...The data of this study indicate that this open school has not been able to meet these long term academic objectives.
>
> The results of this study tend to support the concerns expressed by several authors (Friedlander, 1965; Troost, 1973) about the recent trend in American education away from the more structured type of curriculum. Authors who have supported the movement toward open education in this country (Rogers, 1972; Silberman, 1971) have emphasized the importance of understanding the process of education as well as the final products. These authors have suggested that children should be given the opportunity to explore their own interests, and the freedom to pursue those areas of knowledge to which their natural motivation and curiosity will lead them. It is possible that educational practices founded on these principles do not necessarilty lead to mastery of basic skills in the standard content areas. Yet, mastery of the basic skills such as reading may be even more essential to the child who is attempting to work independently than they are to the child who is working in a more fixed learning environment. (pp. 462-463)

Making Recommendations

Many, if not all, investigators use the conclusions as a basis for making recommendations regarding the use of the information provided by the study or for discussing its implications. The researcher has an opportunity in this section to discuss the meaning of the findings and how the findings and the conclusions may be used by the reader; in other words, here is where the investigator explains the practical uses that may be made of the knowledge he has contributed through his study. He can make recommendations for future practices in the classroom or in business firms, but the recommendations must have as a base the findings of the study.

However, the conclusions and the recommendations should not be confused. A conclusion such as the hypothetical one given above, that Procedure A is likely to produce better results in the classroom than Procedure B, is often incorrectly stated this way:

> Procedure A should be used rather than Procedure B since the students using Procedure A showed significantly greater progress than did those taught with Procedure B.

Worded that way, the statement is a **recommendation** that might be made on the basis of the findings; it is not a conclusion.

Most research studies seem to raise more questions than they answer, and these questions lead to recommendations for further study. These may be simple questions and the research ideas generated may simply be enumerated, or they may be presented in a discussion if the ideas lead to a more complex research study.

Generally speaking, the recommendations and the implications sections allow the writer some freedom in his writing, allow him to speculate a bit, and allow him to use his creative ability in pointing out the possible uses of the findings and in raising further questions for investigation. However, all educational implications, for example, recommendations regarding teaching practices, must be firmly grounded in the findings of the study.

Compiling Supplementary Pages

Usually a complete research report will include some supplementary sections such as an abstract, a bibliography, and one or more appendixes.

The **abstract** is a synopsis of the study and contains a minimum of detail. A succinct statement of the problem (sometimes omitting the detailed questions comprising the subproblems) and a brief description of the procedures are generally given in the abstract. Because the main emphasis in the abstract is usually on the findings, only enough information need be given about the procedures to enable the reader to evaluate whether or not the findings and conclusions are logical. Only the main findings need be given, and the same is true regarding the conclusions. If the writer has sufficient space (most abstracts seem to be about 500 or 600 words long), the researcher may include some of the recommendations; but the procedures and the findings are usually given the most emphasis.

The **bibliography** should be complete, of course, and should include all the studies and periodical articles referred to anywhere in the report. Books in which any of the statistical techniques were delineated should also be included in the bibliography. The bibliography may also contain books, dissertations, or articles not actually quoted or referred to in the report but that were pertinent to and consulted in the preparation of the study. This is not to say that any and all items read in the course of doing the study should be included in the bibliography; it should not be padded with anything and everything remotely related to the topic under study, although this can be a temptation.

The **appendix** should include any special materials used in the research, any questionnaires or other data-gathering instruments, any correspondence granting permission to use certain materials or to conduct the study in given schools or businesses, and **all** raw scores in experimental or prediction (correlational) studies. The raw scores are included in case a reader wishes to analyze the data using a different statistical technique or to check the accuracy of any calculations. The appendix section should also include any large detail tables containing data used in construction of the tables in the body of the study. If the study was an experiment, any procedures manuals prepared for the participants should also be included. The material in the appendix should supply the reader with supplementary details necessary to the conduct of a good study but that need not be presented in the text of the report. All material in the appendix should be referred to in the report and usually should be arranged in the same order in which reference is made to them in the text. Thus, any supplementary material that might be of interest to a reader but that is not essential in the body of the report should be made part of the appendix.

Writing and Organizing the Report

Generally a research report is written in formal style; that is, in third person singular and simple past tense. The writing should be concise, simple, nonpedantic, and **precise**. The following statement, which may seem to follow these criteria, raises several questions: "Each teacher taught a portion of the sections by both the experimental and control methods." What does that writer mean by "a portion"? Does he mean some of the students in each of several sections, or does he mean that the teacher taught some complete sections by experimental methods and others by a traditional method?

Each word must be used properly and must convey exactly the meaning intended by the writer. Consider the following statement: "The data were collectivized according to . . . " The writer obviously meant "categorized," not "collectivized."

All unnecessary words should be omitted. Nonfunctioning words indicate undisciplined writing and reflect imprecise thinking. "It is," "there are," and "which are" are common examples of nonfunctioning expressions in which **it, there,** and **which** have no antecedent. Compare the original and the revised sentences given below:

Original: The writings were scrutinized and comparisons made to establish their validity, and on that basis it was therefore possible to eliminate the recent history of high school beginning bookkeeping enrollments. (30 words)
Revised: After establishing the validity of the writings, the decision was made to eliminate the recent history of beginning bookkeeping enrollments. (20 words)

Original: An investigation should be made of the comparison of business education in the junior colleges and business colleges. (18 words)
Revised: Business education in junior colleges and business colleges should be compared. (11 words)

Original: We might generalize from these early studies and say that the results of early experimental studies have shown... (18 words)
Revised: Results of early experimental studies have shown... (7 words)

Original: It is the tendency on the part of many secondary school principals to rate business teachers... (16 words)
Revised: The tendency of many secondary school principals is to rate business teachers... (12 words)

Original table title: Type of Employment Other than Teaching in which the Business Graduates of the ______ State Teachers College are Pursuing (19 words)
Revised: Nonteaching Employment Pursued by Business Graduates of ______ State Teachers College (11 words)

Perhaps the easiest way to eliminate the nonfunctioning words is to write what one might think are complete and concise sentences and put away the writing for a day or two. Then take it out again, reread it and eliminate all excess words. Modifiers, for instance, are a good place to start the deleting process, and many times **that** or **which** may be eliminated—cut the sentences to the "bare bones." Notice in the above examples how the extra words tend to obscure the meaning of the sentences and how the reader is required to exert considerable effort in determining the real message. A writer should be economical, even parsimonious, with his words in a research report. If a two-syllable word says the same thing a five-syllable one does, opt for the two syllables. Don't hide a good study behind a facade of jargon.

The writer should also refrain from editorializing; he merely reports the findings as facts and makes no evaluative comment about them. He does not intimate his feelings regarding either previously done research or in the recommendations he makes (note the word **must** in the third example below):

It is interesting to note that of the 25, nine were... [**Comment:** Who thinks "it" is interesting besides the writer?]

...has been expressed by "Mr. Republican" himself, the late Senator Robert A. Taft. [No comment necessary.]

> Curriculum planners should keep in mind that a bookkeeping curriculum opens opportunities... Moreover, curriculum planners must remember that the bookkeeping curriculum provides a strong background... [**Comment:** Do the findings really indicate "must"?]
>
> There are three main factors involved in marketing operations. These are goods, people, and operations... Of the three, people are the most interesting. [**Comment:** Perhaps that is true for the writer, but some researchers would rather study operations than people.]

The organization of each section of the report helps the reader also. Each section should begin with a general statement of the topic. Such a general statement should precede any center head, and, in turn, a side head should not directly follow a centered head without an introductory sentence or two between them. The general statement is somewhat like a synopsis placed at the beginning of the section, which is then developed in detail and is followed by a general summary statement. Such an organizational structure permits a reader scanning the report to grasp in general the message without having to "dig it out." This organizational pattern follows somewhat the advice given to a young minister by a master preacher: "Tell them what you are going to tell them, tell them, and then tell them what you told them."

A seemingly elementary suggestion about organizing the report, but one that graduate students tend to avoid, is for the writer, after he has prepared a chapter, to make a detailed outline of what he has written. Such an outline is a good test of the organization of the chapter and will help the writer evaluate the appropriateness of the headings, whether they have parallel wording, and whether an appropriate number of headings have been used. Of course, each section of the report should be written from an outline; but the second outline tells the writer how closely he has followed his original plan and gives him a final check on the overall organization of the report.

Finally, construct a good **title** for the study. The "rules" of good writing apply to the title as well as to the text of the report. Eliminate unnecessary words, but keep those words that make the title descriptive of the study. The title need not, for instance, include such words as "An Analysis of...," or "An Experiment to..." or "A Survey of..." if the rest of it is well worded. Recalling the kinds of conclusions that are legitimately drawn from the different kinds of studies, what kind of a study might have this title: "The Effectiveness of a Noncomputer Business Game in Teaching Post High School Business Problems Classes"? If the title read "An Experimental Study to Determine the Effectiveness..." would you know any more than you do with the first form of the title? Including in the title the method(s) being compared to the game would make the original title a little more descriptive than it now stands.

Consider the title "An Exploratory Study of Self-Instructional Text and Lecture Systems of Teaching COBOL Computer Programming," comparing it with this one: "The Effect on Achievement of Self-Instructional Textbooks or Lecture Systems of Teaching COBOL Computer Programming." Which of the two titles is more informative?

On the other hand, some titles are incomplete and do not give enough information:

1. General Business Teachers in Action
2. Learning Economic Competencies in the Kindergarten

Does the first title refer to classroom activities or professional activities? In fact, what does "in action" mean? Does the second title imply that a survey was made of teachers to see if economic concepts can be learned by kindergartners, or does it mean the researcher tried to find out what economic concepts children do learn in kindergarten? Perhaps an experiment was conducted in which some classes were taught simple economic concepts and others not, or perhaps two or more methods of teaching economic concepts to five-year-olds were compared.

Although the title must be complete enough to let the reader know what the researcher actually did, it need not be an abstract of the entire study, as in these titles:

1. An Analysis of the Minimum Electronic Data Processing Employment Requirements in Selected (State) Businesses for the Purpose of Developing a High School Program of Study for Preparing Individuals for Positions in the Field of Electronic Data Processing
2. A Survey of the 19-- to 19-- Graduates of Eleven Colleges and Universities in (State), with Emphasis upon the Graduates of East (State) College, to Determine Some of the Significant Relationships Between Undergraduate Collegiate Business Education and Business Careers

In the first title, the words "An Analysis of" are unnecessary since any researcher analyzes the data collected in the study. If the purpose of the study is spelled out in the study, it need not be included in the title. The title would be clear enough if it read "Minimum Electronic Data Processing Employment Requirements in Selected (State) Businesses." In the second title, what does the researcher mean, "with emphasis upon"? If the researcher was interested in graduates of a particular college, why did he include all the other college graduates? Does this mean he sampled more heavily the graduates from that one school? If so, does that over-representation distort the findings regarding all eleven schools? One needs to read the study to answer those questions, so wouldn't a title similar to this suffice: "The Value of Undergraduate Business Education Courses to the Business Careers of Graduates from Selected (State) Colleges and Universities"?

The title must not be misleading. A few years ago the title of a journal article was "A Typewriting Experiment in Syllabic Intensity." The title implies that the study was concerned with comparing speed and accuracy scores on typing tests at different syllabic intensities, when, in fact, the report describes a large survey of thousands of typing papers all typed from copy with a syllabic intensity of 1.3 syllables per word. No comparison groups were included in the study.

In review, a good research report will contain these parts, although not necessarily in the order in which they are listed here:

I. Introduction
  A. Background of the study
  B. Purpose of and need for the study
  C. Statement of the problem, including questions that form the subproblems
  D. Theoretical base
  E. Definition of terms
  F. Delimitations of the **problem**
  G. Limitations of the study

II. Review of the literature
  A. Synthesis of past research in the field
  B. Evaluation of past research
  C. Framework for the current study

III. Procedures
  A. Description of population and/or samples and their selection
  B. Description of variables and how they are measured
  C. Detailed account of **all** procedures, including any pilot studies
  D. Description of analysis of data

IV. Findings
  A. Description of the results
  B. Explanation of any seeming inconsistencies or discrepancies in the results
  C. Comparison of current findings with those in other studies

V. Summary, Conclusions, and Recommendations (Implications)

VI. Supplementary pages
  A. Bibliography
  B. Appendixes
    1. Instruments used in gathering data
    2. Raw data
    3. Detailed tables of data
    4. Procedures manuals
    5. Correspondence:
      a. Letters of permission to conduct the study
      b. Cover letters for questionnaires
      c. Follow-up letters
      d. Other pertinent correspondence
    6. Abstract (this may be the first page following the title page)

A carefully done study merits a well-written, well-organized report. It is well worth the effort and time it takes!

References

Armstrong, R. D. *A comparison of three approaches to practice sessions devoted to the development of skill in production typewriting.* Unpublished doctoral dissertation, Pennsylvania State University, 1968.

Beattie, A.D. *Relationships between information and attitudes of high school pupils toward personal finance.* Unpublished doctoral dissertation, University of Minnesota, 1962.

Beyda, E. M. *Employee economic education programs by business firms.* Unpublished doctoral dissertation, Indiana University, 1967.

Boak, R. T. & Conklin, R. C. The effect of teachers' levels of interpersonal skills on junior high school students' achievement, *American Educational Research Journal,* 1975, **12**, 537-549.

Borg, W. R. & Gall, M. D. *Educational research, an introduction* (2d ed.). New York: David McKay Company, Inc., 1971.

Boss, M. L. *A micromolar behavioristic approach to dictation skill building in beginning shorthand.* Unpublished doctoral dissertation, Colorado State College, 1967.

Brandmeir, J. W. *An analysis of attitudes of business administration majors at graduated, senior, and freshman levels about selected business system environmental concepts.* Unpublished doctoral dissertation, University of Washington, 1969.

Campbell, D. T. & Stanley, J. C. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally Publishing Company, 1966.

Chaney, L. H. *The development of an instrument for evaluating student teachers of skills subjects in business.* Unpublished doctoral dissertation, The University of Tennessee, 1972.

Cyphert, F. R. & Gant, W. The Delphi technique: A tool for collecting opinions in teacher education. *Journal of Teacher Education,* 1970, **21**, 417-425.

Davidson, D. G. & Kilgore, J. H. A model for evaluating the effectiveness of economic education in primary grades. *The Journal of Economic Education,* 1971, **3**, 17-25.

Dvorak, E. Writing the research report. In C. C. Calhoun & M. C. Hillestad (Eds.), *Contributions of research to business education* (9th Yearbook). Washington, DC: National Business Education Association, 1971.

Ebel, R. L. *Essentials of educational measurement.* Englewood Cliffs, NJ: Prentice-Hall, Inc., 1972.

Erickson, L. W. & Oliverio, M. E. *Evaluative criteria for survey instruments in business education.* Monograph 111, Cincinnati, OH: South-Western Publishing Co., 1964.

Ford, W. S. & Patterson, H. F. *The development and use of a Q-sort for determining the most characteristic problems of student teachers in business education.* Unpublished doctoral dissertation, Colorado State College, 1967.

Galfo, A. J. & Miller, E. *Interpreting educational research* (2d ed.). Dubuque, IA: Wm. C. Brown Company, 1970.

Ganser, C. J. *An evaluation of the understanding of economic concepts by business education and social studies undergraduate teaching majors at Wisconsin State University, Whitewater.* Unpublished doctoral dissertation, The University of Wisconsin, 1969.

Good, C. V. *Essentials of educational research.* New York: Appleton-Century Crofts, 1966.

Gronlund, N. E. *Measurement and evaluation in teaching* (2d ed.). New York: Macmillan Company, 1971.

Hansen, E. D. *A comparison of the teaching behaviors of creative and less creative basic business teachers.* Unpublished doctoral dissertation, University of Minnesota, 1967.

Hartman, W. M. *The preparation and tryout of materials for a combination beginning shorthand-typing class: A pilot study.* Unpublished doctoral dissertation, University of Northern Colorado, 1970.

Hayes, M. I. *Shorthand homework practice under attention-focusing and vocabulary-expanding conditions.* Unpublished doctoral dissertation, University of Northern Colorado, 1971.

Heimerl, B. & Halldorson, M. Data gathering techniques. In C. C. Calhoun & M. C. Hillestad (Eds.), *Contributions of research to business education* (9th Yearbook). Washington, D.C.: National Business Education Association, 1971.

Helmer, O. *Analysis of the future: The Delphi method.* Santa Monica, CA: Rand Corporation, 1967.

Helmstadter, G. C. *Research concepts in human behavior.* New York: Appleton-Century Crofts, 1970.

Hillestad, M. C. *Factors which contribute to the difficulty of shorthand dictation.* Unpublished doctoral dissertation, University of Minnesota, 1960.

Hopkins, C. R. (Ed.) *Needed research in business education,* Research Bulletin No. 3, Delta Pi Epsilon: St. Peter, MN, 1972.

Houston, S. R. & Teglovic, S., Jr. The computer and research. In C. C. Calhoun & M. C. Hillestad (Eds.), *Contributions of research to business education* (9th Yearbook). Washington, DC: National Business Education Association, 1971.

Karaim, W. J. *A comparison of writing practices of two diverse achievement groups of shorthand students when writing from dictation.* Unpublished doctoral dissertation, The University of North Dakota, 1968.

Kennedy, J. J. The use and misuse of analysis of covariance. *NABTE Review,* 1977, **4**, 10-11.

Kerlinger, F. N. *Foundations of behavioral research* (2d ed.). New York: Holt, Rinehart and Winston, Inc., 1973.

Lee, L. W. *Personality characteristics and job satisfactions of certified professional secretaries and non-certified secretaries.* Unpublished doctoral dissertation, University of Northern Colorado, 1970.

Madison, L. *The identification of the need for employees and the general characteristics of office positions of technological nature.* Unpublished doctoral dissertation, Indiana University, 1969.

McLean, G. N. *The effectiveness of model office, cooperative office education, and office procedures courses based on employee satisfaction and satisfactoriness eighteen months after graduation.* Department of Health, Education and Welfare, Project #4-552; Grant #NIE-G-74-0089, June, 1975.

Mouly, G. J. *The science of educational research.* New York: Van Nostrand Reinhold Company, 1970.

Ober, B. S. *An analysis of the business working papers typed by beginning office workers.* Unpublished doctoral dissertation, The Ohio State University, 1974.

Osborn, D. J. *Methods of teaching shorthand symbol reading and the effect of shorthand reading skill on recording dictation.* Unpublished doctoral dissertation, University of Northern Colorado, 1970.

Osgood, C., Suci, G., & Tannenbaum, P. *The measurement of meaning.* Urbana, IL: University of Illinois Press, 1957.

Rabinowitz, W. and Roberts, D. Common Glaws in Research Design, *NABTE Review,* **4, 5-8. 1977.**

Rahe, H. (Ed.) *Index to doctoral dissertations in business education, 1900-1975.* St. Peter, MN: Delta Pi Epsilon, 1975.

Resta, P.E., & Baker, R. L. Components of the educational research proposal. In R. L. Baker & R. E. Schutz (Eds.), *Instructional product research.* New York: American Book Company, 1972.

Sackman, H. *Delphi critique.* Lexington, MA: D. C. Heath and Company, 1975.

Scott, W. A. & Wertheimer, M. *Introduction to psychological research.* New York: John Wiley & Sons, 1962.

Spear, M.E. *Practical charting techniques.* New York: McGraw-Hill Book Company, 1969.

Stephenson, W. *The study of behavior.* Chicago: University of Chicago Press, 1953.

Sticht, T. G. (Ed.) *Reading for working: a functional literacy anthology.* Alexandria, VA: Human Resources Research Organization, 1975.

Sullivan, H. J. Classifying and interpreting educational research studies. In R. L. Baker & R. E. Schutz (Eds.), *Instructional Product Research,* Chapter 1. New York: American Book Company, 1972.

Thorndike, R. L. *Educational measurement* (2d ed.). Washington, DC: The American Council on Education, 1971.

Tuckman, B. W. *Conducting educational research.* New York: Harcourt Brace Jovanovich, Inc., 1972.

United States Employment Service. *Handbook for analyzing jobs.* Washington, DC: Government Printing Office, 1972.

Van Dalen, D. B. *Understanding educational research* (3d ed.). New York: McGraw-Hill, Inc., 1973.

West, L. J. Sampling techniques. *Delta Pi Epsilon Journal,* 1963, 5, 115-125.

Wright, R. J. The affective and cognitive consequences of an open education elementary school. *American Educational Research Journal,* 1975, 12, 449-468.

DELTA PI EPSILON fraternity, founded in 1936, is an honorary graduate organization for men and women devoted to the advancement and professionalization of business education. Through its ideals of service, leadership and cooperation, the fraternity strives to make significant and unique contributions to professional growth and scholarly achievement in business education. In the words of its founder, Dr. Paul Lomax, can be seen the scope of the fraternity, "The professional interests of Delta Pi Epsilon encompass the whole of business education in relation to the entire fields of American business and American education. Its membership... must always think in terms of the common good and advancement of all our business teachers and of all students who pursue courses in business education."

DELTA PI EPSILON CHAPTERS, INSTITUTIONS, CITIES, STATES

ALPHA: New York University, New York, NY, 10003
BETA: Oklahoma State University, Stillwater, OK 74074
GAMMA: University of Pittsburgh, Pittsburgh, PA 15260
DELTA: University of Cincinnati, Cincinnati, OH 45221
EPSILON: Boston University, Boston, MA 02215
ZETA: University of North Carolina, Greensboro, NC 27412
ETA: University of Denver, Denver, CO 80210
THETA: Indiana University, Bloomington, IN 47405
IOTA: Syracuse University, Syracuse, NY 13210
KAPPA: University of Michigan, Ann Arbor, MI 48103
LAMBDA: Northwestern University, Evanston, IL 60201
MU: University of Tennessee, Knoxville, TN 37916
NU: University of Kentucky, Lexington, KY 40506
XI: University of Florida, Gainesville, FL 32603
OMICRON: University of Iowa, Iowa City, IA 52240
PI: Ball State University, Muncie, IN 47306
RHO: Ohio State University, Columbus, OH 43210
SIGMA: University of Oklahoma, Norman, OK 73069
TAU: Columbia University, New York, NY 10027 (inactive)
UPSILON: University of Mississippi, University, MS 38677
PHI: University of Minnesota, Minneapolis, MN 55455
CHI: Pennsylvania State University, State College, PA 16802 (inactive)
PSI: University of Southern California, Los Angeles, CA 90007
OMEGA: George Peabody College for Teachers, Nashville, TN 37203
ALPHA ALPHA: University of Northern Colorado, Greeley, CO 80631
ALPHA BETA: University of Illinois, Urbana, IL 61803
ALPHA GAMMA: University of Houston, Houston, TX 77004
ALPHA DELTA: Emporia State University, Emporia, KS 66801
ALPHA EPSILON: North Texas State University, Denton, TX 76203
ALPHA ZETA: Temple University, Philadelphia, PA 19122
ALPHA ETA: University of Wisconsin—Madison, Madison, WI 53706
ALPHA THETA: University of Texas, Austin, TX 78712
ALPHA IOTA: University of Colorado, Boulder, CO 80304
ALPHA KAPPA: San Francisco State University, San Francisco, CA 94132
ALPHA LAMBDA: Michigan State University, East Lansing, MI 48823
ALPHA MU: State University of New York, Albany, NY 12203
ALPHA NU: University of North Dakota, Grand Forks, ND 58202
ALPHA XI: Hunter College of the City University of New York, New York, NY 10021
ALPHA OMICRON: University of California at Los Angeles, Los Angeles, CA 90024
ALPHA PI: Wayne State University, Detroit, MI 48202
ALPHA RHO: California State University, Fresno, Fresno, CA 93726
ALPHA SIGMA: Arizona State University, Tempe, AZ 85281
ALPHA TAU: University of Northern Iowa, Cedar Falls, IA 50631
ALPHA UPSILON: University of Nebraska, Lincoln, NE 68508
ALPHA PHI: Northern Illinois University, DeKalb, IL 60115
ALPHA CHI: Rider College, Lawrenceville, NJ 08648
ALPHA PSI: Mankato State University, Mankato, MN 56001
ALPHA OMEGA: Brigham Young University, Provo, UT 84601
BETA ALPHA: Indiana University of Pennsylvania, Indiana, PA 15701
BETA BETA: Southern Illinois University at Edwardsville, Edwardsville, IL 62025
BETA GAMMA: Virginia Polytechnic Institute and State University, Blacksburg, VA 24061
BETA DELTA: University of Georgia, Athens, GA 30601
BETA EPSILON: San Jose State University, San Jose, CA 95192
BETA ZETA: Indiana State University, Terre Haute, IN 47809
BETA ETA: Bowling Green State University, Bowling Green, OH 43403
BETA THETA: University of Wisconsin—Whitewater, Whitewater, WI 53190
BETA IOTA: Illinois State University, Normal, IL 61761
BETA KAPPA: Portland State University, Portland, OR 97207
BETA LAMBDA: Shippensburg State College, Shippensburg, PA 17257
BETA MU: Central Connecticut State College, New Britain, CT 06050
BETA NU: Utah State University, Logan, UT 84321
BETA XI: Memphis State University, Memphis, TN 38111
BETA OMICRON: Southern Illinois University at Carbondale, Carbondale, IL 62901
BETA PI: California State University, Los Angeles, Los Angeles, CA 90032
BETA RHO: Western Michigan University, Kalamazoo, MI 49001
BETA SIGMA: University of Wisconsin—Eau Claire, Eau Claire, WI 54701
BETA TAU: Georgia State University, Atlanta, GA 30303
BETA UPSILON: Pittsburg State University, Pittsburg, KS 66762
BETA PHI: Montclair State College, Upper Montclair, NJ 07043
BETA CHI: Western Illinois University, Macomb IL 61455
BETA PSI: Eastern Illinois University, Charleston, IL 61920
BETA OMEGA: Louisiana Tech University, Ruston, LA 71270
GAMMA ALPHA: Eastern Michigan University, Ypsilanti, MI 48197
GAMMA BETA: Trenton State College, Trenton, NJ 08625
GAMMA GAMMA: Virginia Commonwealth University, Richmond, VA 23284
GAMMA DELTA: University of Rhode Island, Kingston, RI 02881